

US muffs greenhouse gathering

The United States administration must find a better way of saying what it believes should be done about the prospects of global climatic change. By blowing cold and then hot in Washington last week, it alienated both enemies and friends.

GOVERNMENTS, especially that of the United States, are making too much heavy weather of the greenhouse effect. So much was evident at last week's shambolic conference in Washington. The contrast between the opening and closing speeches by President George Bush has been much remarked on (see page 799). At the outset, he emphasized the need for more research, for "the facts" — otherwise, "the stuff that science is made of" — and the need to strike a balance between environmental and economic considerations. But two days later, his theme was "global stewardship", the concept of this generation's obligations towards its successors which is plainly destined to be the fashion of the 1990s (see *Nature* 344, 179; 15 March 1990). The truth is that, while most of what the United States administration says about the persisting uncertainties about global change and the economic basis of modern society is correct, its message is muffled, even muffed. All this speechmaking is a poor substitute for a coherent policy on climatic change.

Even "the facts" are not nearly as uncertain as last week's conference was told. The reality of the greenhouse effect is not disputed. Of the solar radiation reaching the Earth which is not immediately reflected by clouds and surface ice, some 15 per cent is trapped near the surface by the constituents of the lower atmosphere. If it were otherwise, the climate would be that of the last global glaciation (when atmospheric carbon dioxide was a small fraction of what it is at present). The issue is therefore not whether the greenhouse effect is real, but whether it is possible quantitatively to estimate the consequences of recent changes in the composition of the atmosphere — not merely the additions of carbon dioxide from the burning of fossil fuels, but refrigerant CFCs (chlorofluorocarbons) and methane, which is a substantial by-product of oil exploration and possibly a consequence of the intensive use of fertilizers in agriculture. In these circumstances, the questions that matter are not whether recent additions to the lower atmosphere will change the climate, but when these changes will be significant in the sense of being recognizable, when they will be significant economically and when they will be irreversible (as they would be if the ice-caps should melt).

Thirty years of energetic enquiry (and argument) have not settled these issues, and for reasons which are in no sense mystifying. One notorious difficulty is that regional climate change is a poor guide to what happens globally,

the accurate estimation of which is hampered by the lack of global data. Another is that climate models necessarily take fuller account of direct effects than of the feedback processes to which they give rise, but which (if negative) may partly nullify them.

There are also physical uncertainties to contend with. Estimates of the contribution of the CFCs to global warming, for example, are based on the assumption that the only means by which they are destroyed is interaction with stratospheric ozone; what if the peculiar conditions of the Antarctic spring that lead to the formation of an 'ozone hole' also contribute to their removal from the stratosphere (virtually at ground level during the preceding Antarctic winter)? Uncertainty about the fate of carbon dioxide transferred from the atmosphere to the oceans is potentially much more important, especially if much of it should finish up as rock. To list these (and other) uncertainties is simple. To pretend that there will come a time when they have all been cleared up (and when no others have emerged) is to misunderstand the process of research. To imply that nothing can be done until all the uncertainties are resolved — the effect of Bush's address of welcome last week — is wrong.

None of this diminishes the importance of the economic sub-theme in last week's speechmaking. Economics (more strictly, Adam Smith's phrase "political economy"), for all its derided failures, remains the only known way of quantifying people's present needs, their hopes for the future and the trade-off between the present and the future (which is loosely represented by the discount rate). That is why it is essential that the costs of desired abate-ments of identified environmental nuisances should be compared with the costs of other desired public goods — better public education, for example. (The example is not idly chosen: which environmentalist can be sure that an immediate ban on CFCs is preferable to a more understanding public ten years hence?)

Bush edged in this direction last week, but with the blend of enthusiasm and half-conviction he has made his style. It would have been better if he had come clean and openly acknowledged that the uncertainties of estimating greenhouse global warming pale into insignificance in comparison with those of knowing how to get from where we are to a world in which it may be necessary not merely to ban CFCs (worth some thousands of millions of dollars a year) but also to restrict carbon dioxide emissions (the



by-products of industries worth three orders of magnitude more).

That is why Bush's line, for all its apparent reasonableness, smacks of humbug. The urgent question is not so much to know what will be the future course of the Earth's surface temperature if things continue as at present, but to come to grips with the problems that will arise if it is necessary to restrict the emission of all greenhouse gases. Even CFCs are not a simple problem. (China, for one, has not signed the Montreal Protocol.) The difficulties of restricting the emission of carbon dioxide would be three orders of magnitude greater — and will be intrinsically international. The cost may be so great that US taxpayers discover that they cannot afford both compliance and the goal of keeping Los Angeles (and a dozen other cities) free from smog, as required by the Senate's version of the Clean Air Bill that Bush praised last week. (Piquantly, the Senate bill, likely to be echoed by the House of Representatives, requires that gasoline should in future contain a proportion of alcohol derived from corn, which could accentuate the greenhouse effect if more fertilizer means more methane.) The negotiation of international quotas will be a much more serious problem.

That is why the negotiation of a draft convention on greenhouse gases should be the first task of governments. Bush said last week, in his second speech, that the United States will "encourage a convention". Can it not do more than that? Why not "insist on" a convention, even "demand" one? To be sure, as things are, the only appropriate convention there could be would be a document in which the vital numbers defining the points at which various restrictions come into play are left blank, to be filled in later. But that is not a fatal objection. While there are serious problems of equity to be decided, notably those between rich and poor countries, the process of negotiation is likely to take far longer than for research programmes now under way to sharpen estimates of future surface temperature to the point at which they are convincing. □

To European unity?

French and German zeal for a united Europe should prudently concentrate on institution-building.

THE late Sir Winston Churchill's futile day-trip to Paris in 1940 to offer the then-defeated French government political union with the offshore islands to the north was a quixotic gesture born of a sense of danger. So, too, is the declaration last week by President François Mitterrand of France and Chancellor Helmut Kohl of West Germany. What they say is that the time has come to accelerate the pace of change towards political unity within the framework of the European Communities, citing 1993 as a target date. The panic stems from French disquiet about the impending *Anschluss* of the two parts of Germany. The content of the declaration is a measure of German

eagerness to reassure. But, by making light of serious issues staring European people in the face, the dash for unity by 1993 cheapens much of what has been accomplished in Europe in the past three decades and belittles much of what lies in store.

The most serious error underlying the Mitterrand-Kohl declaration is its citation of recent changes in Eastern Europe as the chief impetus for political union within the framework of the European Communities. The same changes might just as well have been taken as the occasion for delay. Why, it might reasonably be asked, settle now for politically united European Communities when the future relationship remains to be determined between the 12 self-satisfied members and the half-a-dozen countries of Eastern Europe (counting is difficult because of the impending *Anschluss* and because Yugoslavia may count for six), not to mention Scandinavia and even Soviet republics such as Lithuania?

The ideal would be a pan-European security agreement within the framework of the Helsinki agreements of 1978. Some hint of the way that wind can be made to blow should emerge from the European Security Conference at Vienna in June. Decisions such as that of the United States to abandon the improvement of the short-range Lance nuclear missile should help. But only after Vienna will it be feasible to talk of common defence policies and armed forces for the European Communities, as Mitterrand and Kohl were doing last week. The difficulty for the European Communities is that the body is a customs union which hopes to be a single economic market three years from now, but which for the time being lacks both the machinery of unified government and the will to create it.

But all is not lost for Mitterrand and Kohl. The problems and opportunities presented to Western Europe by the changes to the East are more interesting than what would have to be, for now, a paper declaration of unity. Take, for example, Eastern Europe's vast investment in scholarship and research, hamstrung though it is by bureaucracy and the lack of funds (see *Nature* 344, 599–620; 12 April 1990). Given impending economic upheavals in the East, there is a danger that the investment will be largely wasted. How should Western Europe respond to that prospect?

The simple answer is that while Western university and research systems remain nationally idiosyncratic, they will make the deals that suit them individually with similar institutions in the East. Nobody in the West would now have it otherwise. But if the process of Western integration had gone further, it would have been possible to give the present process of partnership with the East a greater sense of coherence. And many good projects and institutions would be helped in ways that now seem unlikely to materialize. Yet the European Commission does not have legal authority to intervene in higher education and research, and may not be the most appropriate body to do the job. But if not, what else? Those are the questions on which enthusiasts for European unity should seek to cut their teeth. □

Diplomatic squalls spoil US climate conference

■ White House admits errors

■ European delegations protest

Washington

Disagreements over the proper response to the threat of global changes in climate disrupted last week's White House international environmental conference, causing the United States to withdraw its advertised proposals for a new research collaboration.

Expectations had been low for the hurriedly called conference, but few could have predicted how divisive it would actually become. President George Bush, who announced the meeting four months ago at the Malta summit, said he intended it to emphasize the importance of economics and international research collaboration in the environmental debate. But many of the 18 visiting delegations had been puzzled as to what the United States actually hoped to accomplish (see *Nature* 344, 694; 19 April 1990).

Despite the warning signs, the speed at which the conference decayed into protests and splinter groups apparently took US officials completely by surprise. By the time the meeting ended on 18 April, the United States had retracted or disowned two position papers, quelled a walk-out threat and tossed most of its promised "concrete proposals" overboard to save the foundering summit. One White House official described the two-day conference as a "major embarrassment", an appraisal that found little argument among US environmental groups.

According to diplomatic sources, the adversarial tone of the meeting was first set by John Sununu, the White House chief of staff, the staunchly conservative force behind the White House position that it is too soon to begin measures to counter global warming.

Sununu is said to have caught wind of a press conference planned by the European Communities to criticize the US emphasis on research rather than action. Through an intermediary, he sent word that future global warming negotiations could be seriously damaged by such a move.

As late as a week before the conference, the White House had been planning to assert that there is no scientific evidence for global warming, according to one administration source. However, last-minute lobbying by the White House Office of Science and Technology Policy for a more balanced view seems to have prevailed. In his opening speech, Bush recalled a recent television talk show in which one scientist argued that the Earth could warm

by nine degrees by the end of the next century, while another scientist "saw no evidence of rapid change".

"Two scientists, two diametrically opposed points of view. Now where does that leave us?", Bush asked. "What we need are facts, the stuff that science is made of." Several delegations had warned the United States against emphasizing its own proposals during the two-day conference, especially since all those attending had agreed that the United Nations' Intergovernmental Panel on Climate Change (IPCC) was the proper body to set international global warming policy. But at the end of one of the working sessions, the delegates were handed a 12-page US 'Charter for Cooperation' including four specific proposals, apparently intended for approval at the conference.

In the document, the United States laid out plans for an undisclosed number of 'international institutes' for policy-orientated global change research. By combining science and economics, the centres could "provide a bridge between scientific research and the policy process". Other proposals included a computer network for global change research and international agreements for data sharing.

European delegates immediately objected to the unexpected document, and the United States withdrew it, claiming that it had been mistakenly released. "It was a tactical error", said White House science adviser D. Allan Bromley, the conference's co-chairman.

Michael Deland, chairman of the White House Council on Environmental Quality, told reporters: "Sometimes in the confusion of a conference, things get passed out that aren't supposed to be." But an hour later, US officials were defusing yet another crisis.

According to the environmentalists who discovered it, a two-page list of "talking points" and "debates to avoid" was accidentally left on a podium after an administration press conference. The internal memorandum, quickly leaked to

reporters, advised US officials which issues to emphasize and which to sidestep.

It is "not beneficial to discuss whether there is or is not warming, or how much or how little warming. In the eyes of the public we will lose this debate", the primer suggested. "A better approach is to raise the many uncertainties that need to be better understood on this issue." When speaking to the press, "don't let reporters position this conference as an attempt to delay serious decisions on this issue", the memorandum advised.

Bromley confirmed that the memorandum was authentic, but said that it had been produced by a "low-level White House employee" and had not dictated his responses.

As disastrous as the meeting was for environmental diplomacy, it may have actually achieved some of its goals in bringing economics to the global warming debate table. Several delegates said that they agreed with the US position that better economic analysis of global warming countermeasures is necessary.

Bromley said that the "gratifying news" from the conference was that many countries, including France, Canada and West Germany, were "systematically pulling back from their commitments" to targets such as 20 per cent reductions in carbon emissions by 2005. "They hadn't thought about the economics before", he claimed.

Bromley may be partly right. Although the French, whose heavy use of nuclear power places them among the lowest producers of carbon, have never moved very strongly for domestic carbon targets, carbon-target initiatives in both Canada and West Germany have run into unexpected political problems recently.

In both cases, the cost of the measures to the domestic economy has been a stumbling point. The West German initiative has been sidetracked by concern over the economic and environmental burden of East Germany. And a Canadian 'green paper' (consultative document) on carbon dioxide targets, which was expected early this spring, has been delayed for further economic analysis.

But several delegates said they were offended by the suggestion that they had come to learn economics at the United States' knee. "Do you believe we could go home and take such measures without anybody asking 'What are the economic effects?'", J. G. M. Alders, environmental minister for the Netherlands, asked rhetorically at a press conference. Those present also took issue with US statements that attributed political delays in carbon dioxide targets to "the Europeans learning about economics". As one participant put it, claiming that an abortive two-day meeting raised anyone's economic consciousness smacks of "searching through the conference debris for a success".

G. Christopher Anderson

EMBRYO RESEARCH

Britain votes yes

London

As *Nature* went to press, British Members of Parliament decided by 364 votes to 193 to allow carefully regulated research on human embryos up to 14 days from conception. □

Market machinations

Washington

THE race to market a genetically engineered monoclonal antibody (mAb) for the treatment of Gram-negative bacterial infections took an unexpected turn last week when Xoma Corporation filed for infringement of its newly issued patent by its rival biotechnology company, Centocor Corporation.

At the same time as the US Patent and Trademark Office issued a patent, Xoma, based in California, filed a declaratory judgement against Centocor of Philadelphia, asking a San Francisco court to declare that the marketing of Centocor's mAb, Centoxin, would infringe Xoma's patent.

This latest action by Xoma could spark off the kind of protracted legal battles already seen over conflicting patent rights to the blood-clot dissolving agent tissue plasminogen activator, and erythropoietin, used to treat anaemia in patients with kidney failure. The broad patent issued to Xoma covers a "method for treating humans with Gram-negative sepsis and/or septic shock using anti-endotoxin mAbs from any mammalian source, including mouse or human", says Carol deGuzman of Xoma.

In the United States, about 250,000 patients are affected by Gram-negative sepsis each year, 80,000–100,000 of whom die from septic shock. Hospitalized patients recovering from surgery are at greatest risk.

This is regarded as the first major therapeutic use of mAbs, and both Xoma and Centocor are awaiting approval of their respective products by the US Food and Drug Administration (FDA). Xoma's product, Xomen-E5, and Centocor's Centoxin represent an important short-term earnings potential for both companies. At stake is an annual US market projected at \$350–400 million.

Responding to last week's court action by Xoma, Charles Cabot, of Centocor says it is without merit and that it "will not hold us [Centocor] off the market, or delay our product coming to market by a single day".

Although Xoma filed for FDA approval of Xomen-E5 in March 1989, seven months ahead of Centocor's Centoxin, stock analyst Peter Drake believes "the FDA will address both applications roughly concurrently", with Xoma having no more than "a few months lead time with the commercialization of its product — if at all".

Both Xoma and Centocor will receive a series of patents, and ultimately the two companies will reach some form of cross-licensing agreement as they commercialize their respective products, says Drake.

Until now, much of the controversy between the two companies had focused upon clinical trial data. Xomen-E5, which Xoma is marketing in conjunction with Pfizer Pharmaceuticals, is a mouse-derived mAb, whereas Centoxin is a human mAb. Cabot points out that in phase III trials, no human antibody response was elicited against the drug Centoxin, and there were no side effects.

This represents a "strong safety profile for any pharmaceutical product".

Although safety and efficacy are important criteria for market acceptance of a product, Centocor may be unduly handicapped by its lack of marketing muscle. The question remains as to whether Centocor, which says it will retain marketing rights to Centoxin, can build up a hospital-based sales effort that will allow it to compete effectively with Pfizer.

Centocor has 20 days in which to respond to the court action. **Diane Gershon**

Amgen plays for time

Washington

THE convoluted saga over US rights to erythropoietin (EPO) twisted in Amgen's favour last week when a Washington federal appeals court upheld a request by Amgen to stay the injunction that would have been imposed had Amgen failed to tender a temporary cross-licensing agreement with its rival, Genetics Institute. In short, the court's decision has removed Amgen's incentive to comply with an earlier district court order issued by Judge Young in Boston on 14 March, which would have compelled both Amgen and Genetics Institute to cross-license their respective EPO products, until the patent dispute surrounding the original patent ruling of 11 December 1989 is resolved (see *Nature* 344, 278; 22 March 1990).

Stock analyst Peter Drake sees this as "a very smart and clever strategy", and hails it as "an important and meaningful victory for Amgen in this ongoing battle". As Melinda Lindquist of Genetics Institute points out, "it removes the 'stick' that Judge Young had to compel the parties to comply. Genetics Institute had already complied with his request, so the stick was to be used against Amgen".

For the first time in this patent dispute, Genetics Institute would seem to have been outmanoeuvred by Amgen. "Amgen has further delayed the approval of Marogen through this particular court action", says Drake, which allows Amgen's EPO product, Epogen, to become more entrenched in the US marketplace. In his view, the reason that Amgen wants to delay is that the "US Food and Drug Administration (FDA) will not approve Marogen [Genetics Institute's EPO product] until the two companies have come to a resolution of this particular patent problem".

Once again, Genetics Institute has been thwarted in its attempt to find a legal loophole around the 'orphan-drug' status awarded to Epogen, which provides Amgen with seven years of marketing exclusivity in the United States. FDA could still approve Marogen without the court's backing if, as Genetics Institute

claims, Marogen is found to be a different product due to differences in glycosylation. Or the FDA could remove Epogen's orphan-drug status, because by manufacturing and selling Epogen in the United States, Amgen is in infringement of Genetics Institute's patent.

In addition, Genetics Institute argues that the number of kidney dialysis patients who could be treated with EPO is above the 200,000 level for orphan-drug status eligibility.

In order to speed the resolution of this seemingly never-ending patent dispute, the appeals court has agreed to expedite reviews of the December court ruling, which upheld the central claims of both companies' US EPO patents, while considering both patents partially invalid and mutually infringing (see *Nature* 342, 846; 1989). Although the courtroom confrontation is set to continue, no dates have been set for the appeals.

Diane Gershon

UK RESEARCH COUNCILS

Advisory board in place at last

London

THE independent membership of the reconstituted Advisory Board for the Research Councils (ABRC) was announced last week, nearly three weeks after its planned start-up date of 1 April (see *Nature* 344, 12 April 581, 1990).

Sir Eric Ash (rector of Imperial College, London), Sir Charles Reece (a member of the Universities Funding Council) and John Flemming (executive director of the Bank of England) survive from the last incarnation of the ABRC. The three newcomers are Professor Richard Gardner (director of the Imperial Cancer Research Fund's Developmental Biology Unit in Oxford), Professor Michael Hart (from the Physics Department at the University of Manchester) and Professor Ian Shanks (chief scientist with Thorn EMI and visiting professor at the University of Glasgow).

Peter Aldhous

United States still angry

Washington & Paris

In an unusually frank report, the US State Department last week reiterated its decision not to rejoin the United Nations Educational, Scientific and Cultural Organization (UNESCO).

Since the United States pulled out of UNESCO in 1984, the organization has made little progress towards needed reforms, the report says. UNESCO remains badly managed, refuses to reduce its bloated bureaucracy, and continues actively to support the Palestinian Liberation Organization and programmes that could result in media censorship.

"Bluntly stated, UNESCO needs the United States as a member far more than the United States needs UNESCO", the report states. In declining to add an estimated \$50 million a year in dues to the \$189 million UNESCO annual budget, "the leverage we retain as a sought-after non-member in some instances is greater than we would wield simply by being 1 vote among 161 others". The State Department paid particular attention to UNESCO's performance in the two years since Federico Mayor of Spain took over as director-general from Amadou-Mahtar M'Bow of Senegal. Despite Mayor's promises of substantial reforms, the report says problems in the size and efficiency of UNESCO's bureaucracy and its lack of budgetary restraint have not been addressed. Both the budget and staff have continued to grow.

Mayor has a reputation as a "well-intentioned but poor administrator", the report says. In recent months, it claims, much of the UNESCO staff has openly rebelled against Mayor's attempts to restructure the bureaucracy. It quotes a memorandum from one UNESCO delegate who complains that "the wishes and intentions of the member states... are either being misunderstood, ignored, side-tracked, or flouted". Although Mayor had previously promised to end the New World Information and Communication Order, a controversial initiative that sought to ensure "balance" in media reporting, he has continued to support the programme internally, the State Department asserts. The initiative is seen by the State Department as a device to censor the press, a concern that was raised again this year when Mayor selected a Soviet national to run UNESCO's newly created communications sector.

Mayor has also been unable to control the size of his staff or the use of contractors, the report claims. Last year, UNESCO employed 550 consultants, a number equal to nearly a quarter of its staff. And despite promising that he would cut staff by 700 people, Mayor has

actually increased staff size, according to the State Department study. More than 70 per cent of UNESCO's budget is spent in its Paris headquarters, rather than in the field.

Interviewed in Washington last week, Mayor lashed out at the US report as "biased" and "politically motivated". The study ignores many of the positive changes in UNESCO such as its new literacy initiative, he said. Postponing US re-entry to UNESCO, he added, will "make the implementation of UNESCO's reform process more difficult".

He also contested the claim that the United States can exert more effective leverage from outside. Other UNESCO officials go further, saying they believe that the United States has already done all the damage it can do. They point out that many US observers, such as Representative Jim Leach (Republican, Iowa), believe that the United States could even be embarrassed by staying outside UNESCO. "It was nuts to get out and it is nuttier still not to rejoin", Leach said in a congressional statement last week.

Some UNESCO officials say that many of Mayor's difficulties have come from trying to satisfy US concerns. The recent staff work stoppage that the State Department report mentions was due to US-endorsed austerity measures rather than clumsy management, says Yaw Turkson, a Mayor aide. Mayor "wants to change, but the staff said 'no'. So the [US] report comes back and says the staff is against him", he says.

Turkson says that the real victims of a continued US boycott are the developing countries. By withholding its funds to force UNESCO reforms, the United States is trading literacy in developing countries for its own diplomatic aims, he says. "Even if it's true that the United States has more clout out than in, how can they publicly say it? Have they no interest in the Third World?", he asks.

He also points out that the US report stands in marked contrast to a recent report from the UK House of Commons Select Committee on Foreign Affairs, chaired by Timothy Sainsbury (see *Nature* 344, 368; 1990). Although Sainsbury reaffirmed that Britain was still not ready to rejoin UNESCO, the UK report noted that improvements had been made within the organization and the situation should be appraised again in another year.

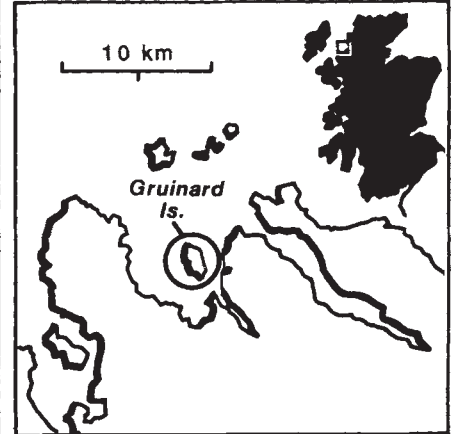
Mayor said he will ask the State Department for a further report. He did not specify when he would make the request, but he said that UNESCO would fully cooperate with US observers and information requests.

G. Christopher Anderson & Peter Coles

Gruinard Island handed back

London

THE UK Ministry of Defence is handing back a remote Scottish island, nearly 50 years after it was requisitioned for wartime biological warfare experiments. On 1 May,



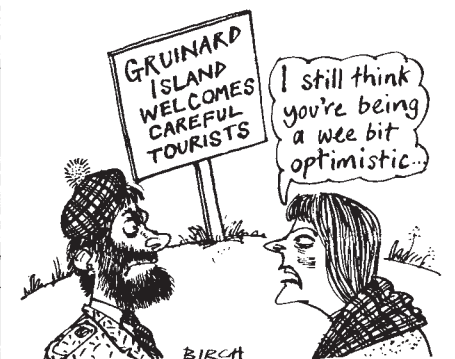
Gruinard Island, off the northwest coast of Scotland will be sold back to the heirs of its original owner, for £500 — equal to the compensation paid to the owner at the end of the Second World War.

Bombs containing anthrax (*Bacillus anthracis*) spores were detonated on the island in 1942 and 1943, as part of a British, Canadian and United States research programme set up in response to fears that the Germans were developing biological weapons. Anthrax was chosen because its resistant spores survive explosions, and are released as an aerosol causing fatal disease if inhaled by man or animals.

For most of the intervening period, the island has laid derelict, strictly off-limits to civilians, in the hope that traces of anthrax would eventually disappear. But in the early 1980s, surveys showed that contamination was limited to the immediate area of the experiments. In 1986 and 1987, these areas were decontaminated by spraying with 5 per cent formaldehyde in sea water.

The island was passed as safe in 1988, by a group of independent scientists led by Professor Bill Stewart, secretary of the Agricultural and Food Research Council, after a local farmer had grazed his flock of 40 sheep on the island for several months with no ill effects.

Peter Aldhous



Star warriors seek new foes

Washington

IN one of the strangest plans ever put forward for the peacetime use of a sophisticated military technology, a US physicist and his Soviet counterpart last week proposed that advanced 'Star Wars' lasers might be used to help rid Africa and the Middle East of swarms of locusts. Despite its novelty, their proposal to blast locusts with airborne lasers has already found preliminary support at high levels from military and science advisers in both the United States and the Soviet Union.

Peter Franken, who works in optical sciences at the University of Arizona, says that the idea to attack locusts was hatched in discussions with his long-time Soviet colleague Vladislav S. Letokhov, a physicist at the Institute of Spectroscopy of the Soviet Academy of Sciences. Both had worked with high-power carbon-dioxide lasers and were apparently looking for possible applications for them when they hit upon the desert locusts.

OCEANOGRAPHY

Cruising for southern solidarity

New Delhi

INDIA'S oceanographic research vessel *Sagar Kanya* last week entered Caribbean waters as part of a Commonwealth research initiative described by India as one of the first examples of South-South cooperation in oceanographic research.

The Caribbean Oceanographic Resources Exploration involves a 45-day cruise by *Sagar Kanya* around 13 Caribbean countries to map the marine resources in their exclusive economic zones. The survey will provide data that will help in the exploitation of marine resources and in the control of pollution and coastal erosion.

The Indian government is providing one-third of the \$2 million cost of the project and free ship time, alongside a 40 per cent contribution from the Commonwealth Fund for Technical Cooperation. The remainder will come from the participating Caribbean countries of Antigua, Bahamas, Barbados, Belize, Dominica, Grenada, Guyana, Jamaica, St Kitts and Nevis, St Lucia, St Vincent and Trinidad and Tobago.

It is the first time that India has provided leadership in oceanographic research to less developed commonwealth countries.

According to K. N. Johry, project coordinator and a scientist in the Indian Council of Scientific and Industrial Research, the Caribbean states wanted assessment of their marine resources to be a South-South affair without any involvement from the great powers.

K. S. Jayaraman

When swarming, desert locusts in Africa and the Middle East cause enormous damage. Locust swarms may contain a thousand million insects, voraciously ravaging vegetation over areas as large as 50 square miles. In bad years, locust swarms are estimated to have caused more than \$1,000 million worth of damage to crops.

Franken and Letokhov's idea, at first, was to train airborne lasers on the locusts and, in Franken's words "literally roast them". The plan sounded good to the physicists because it would use lasers, involve cooperation between the United States and the Soviet Union, and because it would avoid the need for chemical pesticides. Plus, says Franken, "the little corpses would be good fertilizer". Since Franken and Letokhov met at the University of Arizona last week, however, it seems the plan has been modified significantly. Preliminary experiments showed that the lasers could do an adequate job of killing the insects and could even be effective at lower intensity by damaging the locusts' wing membranes, thereby grounding them permanently. But the scheme quickly revealed the same logistical Achilles' heel as the Strategic Defense Initiative (SDI) itself. The team estimates that it would take a hundred laser-armed helicopters more than three months to destroy their entomological enemy — clearly not fast enough to combat the migrating swarms.

Nonetheless, the researchers are undeterred. Now, Franken says, they want to expand their scope to "explore possible uses of the modern military to search, find and destroy" locusts. The key, he believes, will be to find the locusts early, before the massive swarms develop. This, Franken says, can be best accomplished with military reconnaissance satellites. The use of lasers could be feasible in such a situation, according to Franken, but they are also entertaining alternative options, such as airborne, military flame-throwers.

The plan, in its modified and original forms, has received support from surprising quarters. Franken says that US science adviser D. Allan Bromley and SDI chief scientist O'Dean Judd have both expressed interest. Academician Yevgenii P. Velikhov has gone so far as to promise support for the project through the International Foundation for the Survival and Development of Humanity, an organization whose members include representatives from the United States and the Soviet Union (see *Nature* 336, 191; 1988).

Notably lacking in all of the work so far, however, are entomologists. George Popov, who has advised the United Nations and the US Agency for Interna-

tional Development on locust control for decades, says that the locusts are now in a recession period, and no large swarms have developed since 1988. Popov says that some remote sensing is already under way through several international efforts using civilian satellite technology.

Franken acknowledges the current deficiencies in the plan and the many "touchy socio-political questions" that remain — such as what African and Middle Eastern nations might have to say about laser-armed helicopters buzzing about. But Franken says he is optimistic because of the increased interest he perceives in the United States and the Soviet Union to "beat swords into plowshares". "This," he says of his locust-zapping proposal, "is a good place to start."

Seth Shulman

MANNED SPACE FLIGHT

West Germany to buy a ride

Munich

THE Soviet Union has succeeded in enrolling another foreign passenger on the now-routine space flights to its space stations. Following in the tradition of Bulgarian, East German, Mongolian, Cuban, Vietnamese, Syrian and French 'cosmonauts', a West German will blast off with an eight-day Soviet mission to space station Mir in March 1992. An agreement to this effect was signed in Moscow on 18 April by the West German Aerospace Establishment (DLR) and the Soviet organization Licensintorg.

Western cosmonauts do not fly on Mir for nothing. The Soviet Union set a price of DM20 million (about \$12 million), which a DLR spokesman said would be "a good investment" because of the opportunity to do human-tended scientific experiments on board. The West German-Soviet Mir mission will include at least seven experiments in protein crystallization, seed growth and other areas.

An Austrian cosmonaut is scheduled to fly to Mir in the autumn of 1991, but the price is too high for the current Austrian science minister, Erhard Busek, who has said in public that he wished his predecessor had chosen to invest more in Austrian research on the ground.

West German astronauts will also fly on the nine-day D2 spacelab mission with the United States space shuttle in 1992. Much more science will be done on D2 than on Mir, said DLR spokesman Volker Kratzenberg-Annie. The D2 astronaut-scientists will include 90 experiments in materials science, process technology, medicine and biology. Kratzenberg-Annie said that DLR is the only organization to have "concrete collaborations" with the US and Soviet space programmes as well as the European Space Agency. Steven Dickman

AIDS RESEARCH FUNDING

Budget division disputed

Washington

THE long-standing and sometimes heated debate over whether the US government spends too much of its biomedical budget on AIDS continues to split researchers into two almost equally large camps, according to the results of a survey released last week by the US Office of Technology Assessment (OTA). But regardless of their views on the budget shareout, most researchers agree that the assault on AIDS is leading to advances that spill over into other biomedical research fields.

Almost half (48 per cent) of the 148 people who responded to the survey "agreed or strongly agreed" that "too much research funding has been diverted to AIDS research from other fields", while almost as many (44 per cent) "disagreed or strongly disagreed" with the statement. Although the majority of those who strongly agreed with the statement were not receiving funds for AIDS research, more than 40 per cent of AIDS researchers also agreed with it.

Government funding for the study of AIDS has soared since AIDS was first diagnosed in the United States. The total budget now stands at almost \$3,000 million and the National Institutes of Health this year will spend about 10 per cent of its \$7,100 million budget on AIDS-related projects.

Although the pace of budget increases seems set to slow, with the Bush administration requesting only a 7 per cent increase in the AIDS budget for 1991, funding for research on AIDS already exceeds that for heart disease and rivals that for cancer, despite the greater number of deaths from those diseases.

Critics of the high level of AIDS funding have been particularly scathing about the practice adopted by some agencies of actually requesting AIDS-related grant proposals when they feel they are receiving an insufficient number of them. This practice can lead to funds being siphoned away from the pool available for unsolicited grant proposals of higher quality.

Yet the dominant sentiment of the OTA survey is not that AIDS spending should be cut — nearly three-quarters of researchers felt that federal spending on AIDS was about right or even too low. Anthony Fauci, director of the National Institute of Allergy and Infectious Diseases, believes that the message of the OTA report is that more money should be spent on biomedical research in general. "Rather than being concerned about making the slice of the pie for AIDS so big, perhaps the pie itself should be increased", he says.

Fauci also contends that the benefits of AIDS research have "transcended inter-

disciplinary lines", a view endorsed by the OTA survey. More than half its respondents thought that AIDS research had led to substantial advances in fields such as virology, immunology and molecular biology, and more than one-third felt that AIDS research had "contributed substantially" to advances in diagnostics and drug development. On the other hand, genetics, pathology and biochemistry were frequently cited in the survey as having received "little to no" contributions from AIDS research.

At least 40 per cent of survey respon-

TECHNOLOGY TRANSFER

Helping Eastern Europe

London

THE fledgling democracies of Eastern Europe are not able to absorb the massive financial aid packages being prepared by Western nations, according to the annual survey for 1989–90 of the United Nations Economic Commission for Europe (UNECE), released last week. The team of European economists, headed by Aleksander Vacic, a Yugoslav, says that aid should first concentrate on building the basic infrastructure of a competitive market economy, if financial grants and technology transfer to the East are not to be wasted.

There are many similarities between Eastern Europe's economic problems and those of the shattered post-war Western European economies. But a re-run of the 1948 'Marshall plan', when the United States injected cash into Western Europe, would not work, the report argues, because "the certainties of the central planning system have been eroded more quickly than the alternative mechanisms of the market can be put into place".

Ray Oakey, from Heriot-Watt University in Edinburgh, who is studying the diffusion of technology into the Hungarian economy, agrees with the UNECE analysis. He believes that Eastern European research institutes present a pool of cheap but high-quality scientific research labour waiting to be tapped by Western European industry. But the freedom from market pressures in the institutes over the past 40 years presents a problem. Even industrial research contracts, from state-owned manufacturing concerns, have often been little more than a "back door method of subsidizing research institutes", Oakey says. Until market pressures are brought to bear, it is difficult for prospective investors to gauge institutes' potential as centres for market-led research.

Oakey believes that the 'Know How Funds', run jointly by the British Foreign

Office and the Overseas Development Administration, could be used to promote joint ventures between British companies and Eastern European research institutes. These funds are to form the backbone of British assistance to Eastern Europe, but so far have been used mainly for training schemes in the financial sector. British companies may need more encouragement than most to invest in Eastern Europe, because Britain lacks the historical and ethnic links with Eastern Europe of countries such as West Germany.

Despite its cautionary tone, the UNECE report does suggest some areas where immediate financial aid to Eastern Europe is needed. Environmental pollution is a pressing problem (see *Nature* 344, 91 & 610; 8 March & 12 April 1990) and transport and telecommunications systems need updating to Western standards.

Environmental improvement projects have not yet formed a major part of planned multinational aid packages. The European Communities' PHAR programme of aid to Poland and Hungary (which should extend to other East European countries soon), amounting to 300 million ECU in 1990, included no environmental schemes in the first round of projects approved in February.

But the European Communities (EC) environment ministers, meeting in Ireland last weekend, agreed in principle to a specific package of aid to help counter the massive environmental problems in the East. A formal outline for the programme will be drawn up at a meeting between EC and East European ministers in Dublin on 16 June. The transfer of cleaner Western industrial technology should be a major thrust.

EC ministers also suggested an environmental code of conduct for EC-based industries moving into Eastern Europe: industries must not be able to sidestep strict EC pollution controls by exporting pollution to the East.

Peter Aldhous

White House says no

Washington

AFTER five years of bitter debate, controversial amendments to the US animal-welfare regulations have been rejected by the Office of Management and Budget (OMB) because of White House opposition.

The regulations have been the target of a storm of criticism from the scientific community since the US Department of Agriculture (USDA), which oversees the Animal Welfare Act for research laboratories, published its proposals for the new animal care standards two years ago. More than 4,000 researchers have written to USDA to complain that the proposed regulations are unnecessary and that the time and money required to satisfy them could cripple research. Animal-welfare groups, meanwhile, have gathered 2,000 responses to complain that the rules would not go far enough (*Nature* 340, 88; 1989).

In a letter to USDA, OMB last week returned the proposed regulations for further study. "We need a more accurate understanding of the potential costs and effects on biomedical research" that could come as a result of the proposed rules, OMB wrote.

OMB's action was prompted by a letter earlier in the week to USDA from White House science adviser D. Allan Bromley. Bromley asked USDA to rewrite the regulations to allow individual research facilities to decide on their own how to achieve certain animal-welfare conditions. Rather than the strict standards for minimum cage size, exercise time and temperature and light levels for laboratory animals that USDA proposes, Bromley suggests more "flexible" standards based on the health, behaviour and appearance of the animals.

Last month, USDA submitted to OMB a draft of a final rule for the part of the regulation that concerns guinea pigs, hamsters, and rabbits. USDA had planned to seek further public comment on another part of the rules that concerns the "psychological well-being" of non-human primates, as well as the welfare of cats and dogs. In his letter, Bromley said he is concerned that the rules for the smaller animals could set a dangerous precedent for the "even more expensive" rules for the larger ones.

If the regulations had been approved, they would have been published and then become law. But because OMB rejected them, citing Bromley's letter, USDA may have to repeat the entire regulatory process from the beginning, says Gwen Rubinstein of OMB Watch, an independent organization that examines cases of regulatory abuse. "OMB erased three years of USDA work with three paragraphs", she says. USDA spokesman

John Duncan says the agency will decide this week how to respond to the OMB veto. USDA can appeal against the decision, or "do what OMB asks us to do" and begin anew on the rules, he says. If the agency is forced to start again, it could be several more years before the animal-welfare regulations finally become law.

In order to prevent USDA from passing its proposed standards, OMB has used almost every regulatory device at its disposal. In 1988, OMB solicited the opinions of 21 other federal agencies on the USDA rules, a lengthy process that smacked of "disingenuous subterfuge intend to subvert congressional intent through further delay", as USDA put it in an angry letter to OMB.

Later that year, when the other federal agencies had submitted their opinions, USDA lashed out at OMB again. "The comment filed in opposition to the regulations by the State Department, which conducts and funds no research, dramatically illustrates the kind of unhelpful reaction we can expect your [actions] to provoke", USDA wrote.

Former Senator John Melcher (Democrat, Montana), one of the legislators who wrote the Animal Welfare Act amendments in 1985, says he has "never seen

anything so misinformed about the legislative process" as last week's letter from the White House.

In the letter, Bromley suggests that the proposed "engineering-type standards" be replaced with "standards of desired performance" similar to the current guidelines issued by the National Institutes of Health.

USDA "need only verify, upon inspection, that the standard operating procedures are in place and that the facility has not failed to protect animal welfare" he wrote.

But guidelines based on some performance standard were not the intention of Congress, Melcher says. "[USDA] has to write legislation based on the law or ignore it entirely. You can't just keep the *status quo*. We fully intended the regulations would set minimum [and larger] cage sizes." Barbara Rich, executive vice president of the National Association for Biomedical Research, says that conforming to the proposed USDA regulations could cost the biomedical community over \$1,000 million. Many of the changes would not appreciably improve animal welfare, but are merely intended to standardize cages sizes and the like. "The point is to avoid arbitrary decisions without scientific basis. Especially not ones that cost millions of dollars," she says.

G. Christopher Anderson

NUCLEAR POWER

Japan seeks plutonium power for the 1990s

Tokyo

DOES Japan need to import plutonium from Europe in the early 1990s? The Japanese government says yes and is busy rushing through plans to build a ¥20,000 million (\$125 million) armed escort vessel to escort shipments of plutonium from reprocessing plants in Europe. But a report released last week in Tokyo by the Washington-based Nuclear Control Institute (NCI) claims that there is no need for imports before 1998 at the earliest.

The plutonium, derived from spent Japanese nuclear fuel, will be used in various kinds of new reactor including experimental and prototype breeder reactors. According to the government, shipments will be needed in 1992 to meet a shortfall in plutonium supplies.

Both NCI and the government agree that there will be a dramatic increase in Japan's use of plutonium in the 1990s. During the past 12 years Japan has used only about 0.2 tons of plutonium per year to operate the experimental breeder reactor Joyo and the prototype advanced thermal reactor Fugen. NCI estimates that annual demand will go to 0.7 tons from 1992 when the prototype breeder reactor Monju reaches criticality and then to over 1 ton when the demonstration advanced thermal reactor Ohma

comes on line in 1999. Japan's Science and Technology Agency (STA) says, however, that the demand will be some 30 per cent higher than this.

This difference, combined with a lower estimate of plutonium output from Japan's small reprocessing plant at Tokaimura accounts for the different predictions of when Japan's plutonium supplies will run short.

But NCI goes on to make a much more radical suggestion. It points out that the United States and some European countries have abandoned plans for commercial reprocessing. If Japan used low-enriched uranium instead of plutonium in advanced thermal and light water reactors, NCI says that imports and the planned expenditure of ¥1 million million on the construction of a commercial reprocessing plant in northern Japan would be unnecessary.

But Itaru Watanabe, deputy director of the nuclear fuel division of the Science and Technology Agency, says that use of plutonium is fundamental to Japan's energy policy. In the Japanese view, the ability to reprocess plutonium is essential to reduce the very high dependence on foreign source of fuel.

David Swinbanks

One in six jobs to go

London

THE Natural History Museum in London is to shed one in six science jobs in the next three years, according to the Museum's 1990-95 corporate plan, released last Monday.

The director of the museum, Neil Chalmers, says that he intends the plan to "concentrate our research effort into a series of programmes concerned with environmental, human wealth and human health issues", adding that "by doing this we plan to avoid the weakness that can arise from spreading ourselves too thinly over too many areas — trying to be all things to all people".

Behind these words lies the hard truth that 98 per cent of the museum's government grant is swallowed by salaries, and jobs will have to go if the museum is to remain flexible and pursue its reorganization plans. By 1992-93, the current staff of 780 will have been cut by 100.

"Some 40 posts have been saved in 1989-90 through natural wastage and

there will be a further reduction of 50-60 posts by early 1991, by natural wastage and redeployment where possible, supplemented by voluntary and, regrettably, by compulsory early retirement where necessary", writes Chalmers in the introduction to the plan, which also notes that 50 posts have been axed since 1983-84. "If we do not take this action, the Museum will gradually be eroded in a haphazard fashion. Posts would be lost in an unplanned way over the years, with immense damage to the Museum." Chalmers urges the government to provide an increase of £4.4 million over the next five years, over and above the levels indicated by the Office of Arts and Libraries (OAL), which reports to the Arts Minister, Richard Luce. This extra money would ensure that the current government grant keeps pace with inflation. The government grant will be £16.9 million in 1990-91, rising to £19.8 million in 1994-95. But this is based on a fixed annual increase of only 4 per cent, and £21.4 million is a more likely figure for 1994-95.

Researchers fear that if the OAL does not accede to this request, described in the report as "crucial", job losses could be much worse than those threatened. The gloom is deepened by the fact that the OAL will not respond to the corporate plan for some months. Apprehension among researchers is almost tangible: all they can do, they say, is to return to work conscious that one in six will soon receive a tap on the shoulder.

The corporate plan has had the effect of "daubing blood on our doorposts", says museum worker William Lindsay, a member of the National Executive of the Institute of Professionals, Managers and Specialists, one of the three trades unions in the museum.

An earlier draft of the corporate plan shows that it was drawn up as a rearguard action in part of a long battle with OAL, which sees the museum more as a public attraction than a scientific research institution. The draft is much more combative than the final version, promising a marked decline in all aspects of the museum's work if OAL does not increase its aid substantially. It emphasizes the museum's success in its own, independent fundraising: activities that have risen from 17 per cent of the total budget in 1987-88 to 25 per cent in 1989-90, with a target of 30 per cent by 1994-95.

"Self-reliance at this level places the Museum among world leaders. The Smithsonian Institution's Natural History Museum in Washington, for instance, still receives 75 per cent of its funding from Federal and State sources", the draft says. These achievements are mentioned in the final plan, but the draft document makes

no mention of job cuts. Naturally enough, researchers feel that the plan as amended "sells out staff".

More positively, the plan sets out a bold strategy of exhibitions, increased budgets for curation (as distinct from research) and a science policy based on research topics "to develop basic and applied programmes relevant to contemporary needs and issues". Research will now be based around six research groups, concerned respectively with biodiversity, environmental quality, living resources, mineral resources, human health and human origins.

But Chalmers emphasizes that these changes cannot be made with the present staffing levels and tradition of tenure. Chalmers sees flexibility is the key — job losses will provide for 25 short-term 'new blood' fellowships. Although Chalmers admits that aspects of the plan are painful, he hopes that the scientific staff "will realize and accept the challenges and benefits of this new way of working".

Henry Gee

PARTICLE ACCELERATORS

Heavy ions on tap in Darmstadt

Munich

A UNIQUE combination of an accelerator, a synchrotron and a storage ring for heavy ions was inaugurated this week at the Gesellschaft für Schwerionenforschung (Society for Heavy Ion Research or GSI) at Darmstadt. Although many heavy-ion accelerators operate around the world, the GSI apparatus is unique in its ability to store the ions for further study. With this combination, which cost DM275 million (about \$165 million), researchers at GSI hope to create as many as 600 new short-lived isotopes of known elements, says spokesman Günter Siegert.

The new apparatus can accelerate ions as heavy as uranium, to energies as high as 1.3 gigaelectron-volts (GeV) per nucleon, and by stripping all 92 orbital electrons away from the nucleus can probe the structure of the innermost electron shells.

Researchers will also use the device to study nuclear physics; some experiments may have astrophysical relevance, as the GSI apparatus can collide heavy ions onto a variety of targets at energies typical of stellar interiors. GSI has applied for a further DM17 million from the West German Research Ministry (BMFT) to create a tumour treatment centre at the new facility.

In November last year, the GSI synchrotron accelerated argon nuclei to an energy of 1.7 GeV per nucleon, the highest energy yet achieved.

Steven Dickman

BSE

Britain not alone in its predicament?

London

BOVINE spongiform encephalopathy (BSE) may be present at low levels in cattle from several European countries and in the United States, according to Keith Meldrum, chief veterinary officer at the UK agriculture ministry. So far, the disease has been thought to be unique to the United Kingdom and the Irish Republic.

Speaking at the Royal Society in London this week, Meldrum said that some of the Irish cases could not be linked to imported British cattle or feed. "It would not surprise me", he said, if other European Communities states had low levels of BSE infection. In the Netherlands and France cattle were fed meal containing sheep tissues in the early 1980s, so may have come into contact with the agent believed to cause both scrapie in sheep and BSE in cattle.

Meldrum also pointed to an outbreak of transmissible mink encephalopathy (a similar disease) in Wisconsin in the mid-1980s, where mink had been fed exclusively on material from cattle and horses. If the outbreak was linked to diet, then this could indicate a sub-clinical infection of BSE in US cattle, Meldrum suggested.

Meldrum's comments came as the Soviet Union introduced a temporary embargo on imports of British beef and dairy products, because of the BSE problem. The UK Ministry of Agriculture, Fisheries and Food, concerned about the loss of export markets, is anxious to make it clear that BSE may not be an exclusively British disease.

Peter Aldhous

Condition of British science

SIR—Your first leading article on 22 March (*Nature* 344, 275; 1990) was entitled "Bringing research back to life". It was subtitled "Having all but killed it off, the British government is now brooding on how best to restore a once-successful research enterprise to good health". These are remarkable statements, which were curiously unsupported in the body of the article by a single figure or statistic. Let us test these statements against some known facts.

Harry Atkinson and Philippa Rogers of the Science and Engineering Research Council have shown that Britain employs 128,000 researchers in all, supported by a further 155,000 technical staff. Does *Nature* really believe that a total of 283,000 scientific staff represents a science that has been "all but killed off"?

The universities, moreover, have done well over the past ten years. University Grants Committee statistics show that there were 40,246 full time academics in 1976–77, and that this figure had risen to 47,038 by 1986–87. Most of this expansion was in the sciences.

Britons publish a lot of papers. The *Evaluation of National Performance in Basic Research* (ABRC Science Policy Studies, No. 1, 1986) showed that, for 1982, we published 432 papers per million population, compared with 483 or 465 for the United States or Canada respectively, and 295, 279 or 174 for Germany, France or Japan respectively.

The quality of our science remains superb. *Current Contents* of 28 November 1988, 12 December 1988 and 6 February 1989 shows that, of the 300 most cited papers published during 1985–86, only the United States produced more than Britain. Our papers attracted more citations than those of the French, German or Japanese.

Our science is well funded. The *International Comparison of Governmental Funding of Academic and Academically Related Research* (ABRC Science Policy Studies, No. 2, 1986) shows that, as a percentage of gross domestic product, we may not match the West German or French figures, but we out-perform the Americans or the Japanese. Moreover, our non-governmental support for science is increasing fast. The medical charities' support for research has doubled in ten years, and now matches that of the Medical Research Council. Industrial support for academic science has also doubled in the past ten years, as has been acknowledged by both the Association of University Teachers (submission to the Department of Education and Science, 1989) and the Labour Party (*Science for the Citizen*, Jeremy Bray, 1989).

Problems remain, of course. Too many

alpha-related projects go unfunded, but sadly this is an international problem. In the United States, both the National Institute of Health and the National Science Foundation are turning down two-thirds of their *alpha*-rated projects. Science would like to grow at a greater rate than the economy, whatever the country, but obviously it cannot.

We lose too many high quality scientists to the United States but, as I proposed in my pamphlet for the Centre for Policy Studies (*Science Fiction*, 1989), this reflects our universities' employment policies (tenure, fixed salaries, dual funding) rather than our overall funding.

Your leading article admitted that its assessment was "subjective", but if you go around depressing people with unsupported statements, we will all go home in despair and your subjective obituary will become self-fulfilling.

TERENCE KEALEY

*Department of Clinical Biochemistry,
Addenbrooke's Hospital,
Hills Road, Cambridge CB2 2QR, UK*

SIR—Your News columns suggest that everyone is so short of scientists that they are trying to poach them from everyone else. The need may be a real one, but my difficulty is the belief that, were cash available, there would be no problem in increasing the supply.

Surely this belief is wholly unscientific? If the concept of evolutionary adaptation is true, then humans are still adapted to live in conditions of subsistence agriculture. (After all, most of them still do so.) Genetic variance, which produces individuals whose ability to manipulate abstract ideas varies between that of the village idiot and the professor of comparative ambiguity, will always produce a small but constant fraction of the population who are potential scientists. There has been insufficient time over the past few centuries for selection to increase this proportion, although such a change might well happen eventually.

That there now seems to be more of them than before is partly due to the increase in world population and partly to the effect of contiguity. In very small communities, the potential scientific heads will rarely if ever meet those of their own kind, so that inability to 'knock spots off each other' will retard the appearance and growth of scientific ideas. As communities increase in population, and tend to overlap, such contact becomes more and more possible, and scientific habits of thought slowly develop, even if not among the mass of humans. It is no surprise that historians of science insist that the dissemination of ideas is a major cause of the rapid growth of science in the past

two centuries.

When one realizes that personal aptitude as well as genetic influences are necessary to produce good scientific heads, it is hard not to think that the present lack of scientists is due to the fact that we have been scraping the barrel for a long time, and that, however we try, there are few more to come by.

P. C. SMETHURST

*89 Apton Road,
Bishop's Stortford, Herts CM23 3ST, UK*

Romanian minority

SIR—Television viewers in many countries will have witnessed a crowd brutally assault members of the Hungarian minority in the Romanian city of Tirgu Mures (Marosvásárhely) on 20 March.

One of the victims was Dr Mihály Péter, professor of microbiology at the School of Medicine and Pharmacology. Péter (61) was taking part in a peaceful demonstration asking that Hungarian be re-established as a language of education at the university. The use of Hungarian was banned under the Ceausescu regime.

Péter is a popular educator who gained international reputation for developing a hepatitis-B vaccine in the early seventies.

We ask our scientific colleagues to express their concern for Dr Péter.

TIBOR DEÁK

*Department of Microbiology,
University of Horticulture,
Budapest, Hungary*

ERVIN BALÁZS

*Agricultural Biotechnology Center,
Gödöllo, Hungary*

Which came first?

SIR—S. Chandrasekhar (*Nature* 344, 285; 1990) says that the human mind created mathematics, through "speculative thought". He is amazed that this artificial creation turns out to be an excellent predictor of the physical world.

I suggest that he has got it the wrong way round. The physical world created the human brain, through evolution. If the world is mathematical, that fact must show in brain function.

Long before humans walked the Earth, there was a payoff in modelling the world as moving objects in an environment. Projectiles fly in conic curves, predators circle, a straight line is the shortest journey. The brain must surely be wired for mathematics.

A day came when the Greeks took mathematics out of their heads and set it on paper. On that day, predicting the world became a team effort. With the results we see today in physics. Impressive but not surprising.

RICHARD BAKER

*22 Cromwell Road,
Kingston KT2 6RE, UK*

Science in fiction

SIR—The late John Treherne's dismissive review (*Nature* 342, 304–305; 1989) almost put me off reading Carl Djerassi's new novel *Cantor's Dilemma* — “what were Djerassi's motives in writing in such a stomach-turning way about such a ghastly bunch of repellent characters?”, he asks.

I discovered that the novel addresses, in a very effective and powerful way, important contemporary questions about the motivation of scientists. I am certain that it will do great service to the scientific community. Non-scientists, who are often baffled by the conflicts, rivalries, paranoia, human insensitivity, apparent social irresponsibility and even fraud that they think they sometimes perceive in us, will find illumination in this compassionate but unflinching treatment of the way scientists live. It is satisfying in itself as a novel and is a first rate read — even Treherne says “Carl Djerassi can write — there is no doubt about that”.

Part of the skill of the novel is that it starts with a bunch of scientists who are indeed portrayed as others see us — complete with narrow views of life and self-seeking ambitions. But then the human and moral dimensions emerge and (for some) the better sides, as characters are put under pressure from the consequences of the pursuit of research. Who exploits whom in a professor–postdoc relationship? Why do such symbioses (if that is what they are) develop in scientific research when they do not seem to in the humanities? How can a superior know whether fabrication of results is going on in his team? How does an honest woman researcher come to terms with sexism on the one hand and positive discrimination on the other?

These and many other contemporary questions are vividly illustrated with all the gut-wrenching fears and tensions that go with them. Some of the issues foreshadow and indeed might in part have been inspired by the controversy about Stehelin which erupted with the award of the Nobel prize to Varmus and Bishop last year; although the book must have been finished before that time, the problem had been smouldering for years and Djerassi, like Varmus and Bishop, lives in California.

Djerassi brings into the 1990s themes that were first explored in an earlier scientific world by C. P. Snow and William Cooper, but he adds new ones that have become more prominent in our more careerist times. There are villains as well as good guys in the story — though not by any means as many as Treherne suggests. The disappointed and self-seeking Krauss can to some extent be compared with Snow's Nightingale or Cooper's Dibdin, but he has an armoury of manipulative skills for persisting towards his ends that are whole worlds of

sophistication greater. Djerassi's skilful deployment of his characters plays on the latent paranoia that a scientist can experience when he fears having made a career-breaking and public mistake, and sets up the central conflict between Cantor and Stafford, trapping them in a struggle with themselves and each other. It is around this tension that Djerassi weaves the contemporary issues and constructs the highly original plot.

I do hope that your negative review will not prevent this book from being as widely read as it deserves.

D.A. REES

Medical Research Council,
20 Park Crescent,
London W1V 4AL, UK

■ John Treherne reviewed the book upon its publication in the United States (by Doubleday). In Britain, it is published today by Macdonald. □

How 'scientific' is whaling?

SIR—As F. Nagasaki rightly points out (*Nature* 344, 189; 1990), “scientific” whaling can easily develop into an emotive issue where scientific considerations are secondary. However, the arguments advanced in support too readily assume the need to kill whales in order to study them. The very same issue of *Nature* describes non-lethal sampling of whale DNA (344, 238; 1990). Assuming adequate analysis of previously obtained material (for example the 600 southern minke in the Japanese survey 1987–89), it should be practicable to use genetic fingerprinting techniques to perform capture-recapture analysis. Reasonable estimates of total stock size and sexual composition will then be obtained. Assessing the age and maturity of individuals is more problematic. Crude estimates might be possible from observable physical characteristics such as length, relative size and so on of various features and sonar reflections. Over a number of seasons, estimates of abundance and sexual composition will give an indication of whether stocks are increasing or decreasing. Although this will take a period of years, it is prudent to consider carefully the case for whaling of selected species. The suggestion of resumption of commercial whaling within a year does seem indecent haste.

A discussion on the need (or otherwise) to determine age-specific natural mortalities would have been useful to help assess the case for ‘scientific’ whaling. If the research is to be seen as truly altruistic and not as a means of attempting to justify a return to commercial-scale whaling as

soon as possible, it would have been better to commission research from another source, that some critics could not claim was tainted.

CHRIS LLOYD MILLS

50 Sapphire Street,
Cardiff CF2 1PZ, UK

SIR—I was glad to see the case for scientific whaling clearly and explicitly presented by F. Nagasaki. He correctly says “many of those who oppose Japan's whaling operations do so for emotional reasons”. But so what? May we not have and express emotions even if they are “without scientific bases”? Many of us who eat beef and mutton (thereby condoning the killing of cows and sheep) would nevertheless feel uncomfortable shooting foxes or killing and eating the flesh of dogs, horses or monkeys. I like whales, and so — if only for emotional reasons — I would feel happier if all whaling ceased.

RALPH A. LEWIN

Scripps Institution of Oceanography,
University of California, San Diego,
La Jolla, California 92093, USA

Star wars

SIR—The small laser-energized ice-rockets described by G. Christopher Anderson (*Nature* 344, 91; 1990) represent just one of a number of proposed new ways of reducing space-launch costs. US Sandia National Laboratories researchers recently proposed ‘coil guns’. Electromagnetic mass drivers were also advocated earlier by G. K. O'Neill for launching material (more cheaply) from the Moon or an asteroid.

By exploiting ‘Star Wars’ global-positioning, sensor, guidance and steering technology, stretched-out packets of such extraterrestrial material, accelerated by gravity towards Earth, could be made to home in on and impact ‘pusher plates’, for example on initially slow vehicles lifted by ‘air-breathing’ ramjets to points just above the atmosphere. By using shock-absorbers and multiple impacts (as in ‘Orion’-Project nuclear-bomb propulsion), power generation and moderate non-damaging acceleration of even fragile vehicles into orbit should then be possible.

For inter-orbital travel, elastic impacts could perhaps be achieved by utilizing rotating-cable impact material, one end of which would link up with a vehicle at zero relative velocity and swing it through a semi-circle to a higher or lower velocity. Two-way travel with virtually no net energy expenditure may then become feasible.

LOUIS A. P. BALÁZS

Department of Physics,
Purdue University,
West Lafayette, Indiana 47907, USA

High hopes for the Space Telescope

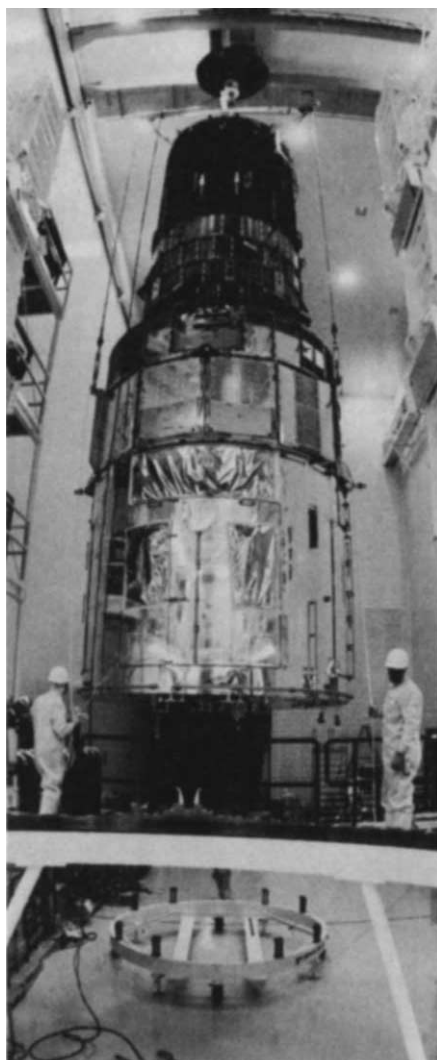
Fred Hoyle

All being well, the Hubble Space Telescope will now be in orbit. The achievement is the culmination of 20 years of politicking and caps a triumph for astronomers over bureaucrats.

THE Hubble Space Telescope is a drama in three acts, each with a staging time of one to two decades. As this article goes to press, the telescope is still on the ground at Kennedy Space Center, the launch having been delayed by a faulty valve in the Space Shuttle's hydraulic system. But when it is in print, we can hope that the first act will be complete with the placing of the telescope in orbit. The second act will be concerned with the scientific results obtained with the telescope, and is expected to last until the early years of the next century, while the third will involve an even greater span of time, for in it the historical impact of the instrument on the development of astronomy in the long term will emerge.

The telescope has what might be termed 'conventional optics', Ritchey-Chrétien, with a primary mirror of diameter 2.4m and a secondary of diameter 31cm, distance about 4.9m from the primary. The Cassegrain focus, where the detection equipment is placed, has a field diameter of 28 arcminutes, of which the central region — a square with a side of about 3 arcminutes — is devoted to the wide field planetary camera. Out to a radius of some 9 arcminutes the rest of the field is divided into four quadrants, each of which is given over to one of the following instruments: the high-speed photometer, the faint-object spectrograph, the high-resolution spectrograph and the faint-object camera (the latter was constructed by the European Space Agency). These five instruments are designed to operate simultaneously, not alternatively as is usually the case for ground-based telescopes. Beyond a radius of 9 arcminutes the field on the focal plane is given over to three fine guidance sensors, two of which are used to control the pointing of the telescope, the third being available for astrometric measurements.

The telescope sits inside a cage, the support systems module, which, as its name implies, provides structural support, thermal control, electrical power from solar arrays, communications and data handling, and the pointing control for the telescope. Although the rate gyroscopes and the reaction wheels employed for the latter purpose may, like the optics, be said to be 'conventional', the measures of precision that have been achieved are anything but conventional. The 2.4m primary is ground and polished to achieve the diffraction limit, about 0.01 arc-



Two years still to go — April 1988, and the Space Telescope is lowered onto a shock-mounted stand in the clean room at Lockheed Missiles and Space Company (NASA).

seconds at the shortest wavelength (1,150 Å) at which the telescope is expected to operate. The same 0.01 arcseconds also defines the pointing accuracy required from the support module. The dynamic range demanded from the instruments is from magnitude +8 on the bright side to +28 on the faint side, the latter corresponding to what is expected for Cepheid variables in the galaxies of the Virgo cloud.

A further precision requirement comes from the plan to replace the first generation of scientific instruments while the telescope is in orbit, replacement being necessary to keep the detection equip-

ment abreast of the state of the art. Such a plan involves the replacements being inserted by astronauts, working from the Space Shuttle, to within a precision of one or two ten-thousandths of an inch.

It is at once clear that the story of the inception and building of so expensive and remarkable an instrument must be unusual. The project was first conceived of in 1946. In the wake of conversations with Leo Goldberg, Lyman Spitzer — then at Yale, before his appointment at Princeton — wrote a report for Rand (a think-tank founded by the Douglas Aircraft Company) entitled "Astronomical Advantages of an Extra-Terrestrial Observatory". Spitzer's concept of a telescope free from atmospheric seeing problems was to remain a gleam in the eye for almost a quarter of a century, until the National Aeronautics and Space Administration had become a dominating organization, penetrating into almost all aspects of science in the United States. In 1969, with their budget steeply in decline, NASA administrators were looking for a programme with strong public appeal to replace Apollo. The choice fell to the Space Shuttle. To add to the rather obvious scientific trivia that were being urged as a case for the shuttle, a project of real depth and magnitude was needed.

The Large Space Telescope, as it was then called, fitted the need, but in an inevitably back-to-front way. Once the support of astronomers in particular and of scientists in general, through the good offices of the US National Academy of Sciences, had been mobilized, and once the unprecedented importance of the Large Space Telescope had been publicly emphasized, it became possible for NASA to argue to Congress that the Space Shuttle was necessary to launch and then service the telescope while in orbit, and also to return it to ground for repairs, adjustments and improvements, so adding to the telescope's lifetime as an instrument of international importance. The reality of course was that it was subservient to the shuttle, not the other way around.

With the support of astronomers and, through the National Academy of Sciences, of scientists generally, it was possible to overcome highly vocal opposition in Congress, to the extent of approval being granted for a design study to be made. After an initial internal struggle within NASA, the Marshall Center at Huntsville, Alabama, was chosen as the

prime contractor. The Goddard Center at Greenbelt, Maryland, which had sought that role, was given the subsidiary one of controlling the design and construction of the ancillary instruments.

By 1972, the Marshall Center had come through with an estimate of \$700 million for the construction of the instrument and for its operation over an expected 15-year lifetime. If the golden rule of two had been applied to this estimate it would have turned out to be not too far off the mark (after allowing for inflation). But NASA's chief administrator insisted that no more than \$300 million could be asked of Congress, thereby undercosting the project by at least a factor of four. It was on this basis that permission for a more extended design study was granted. The work of succeeding years showed the initially projected diameter of 3m for the primary mirror would need to be reduced, if even a remotely acceptable costing was to be achieved. Even a scaling down to 2.4m led the Marshall Center, by 1976–77, to an estimate of \$575 million (for the instrument alone before launch). This was judged too large to secure approval for the start of the manufacturing phase and, as Professor Robert W. Smith remarks in his recent book *The Space Telescope* (see acknowledgement), it seems likely that NASA would have been happy for the project to have been cancelled at this stage — by now the Space Shuttle was up and running on its own. That this did not happen was due to an amazing initiative from astronomers. Setting aside any differences they might have had, they all decided to pull for the telescope. And a selected number decided to lobby Congress directly, proving themselves to be better at the job than NASA had been. Liking the selected number to a rugby scrum, the front row consisted of Lyman Spitzer in the middle, with John Bahcall and George Field as props (all theoreticians, I am happy to point out).

Still more amazingly, those unusual lobbyists succeeded. But, that done, NASA was once again left to manage the construction of the telescope and the astronomers were pushed to the periphery of the project. Over the years to 1983 the ceiling budget was \$575 million, which again would have been almost exactly the right estimate had the golden rule of two been applied to it. The rule states that, after estimating the cost of a new scientific instrument as fairly and as competently as possible, and after adding what seems like a reasonable contingency, you then double the result in order to obtain the real cost of the project. Depend upon it! At \$575 million in 1977–78 the Space Telescope was to be a near perfect example — the correct estimate in 1978 dollars should have been \$1,150 million.

The rule of two arises in part from the buy-in syndrome to which all industrial

consultants are subject, even the most honest and accurate of them. At every uncertainty consultants give optimistic projections, and many small optimistic projections inevitably add up to a large optimistic projection. Because this matter comes as naturally as breathing to consultants, I have never seen any point in attempting to fight it. You cope with it by adding a factor of about 1.3 to the original estimate, which you do your best to keep hidden from your industrial contractors. The remaining two-thirds of the escalation comes from unforeseen technical factors, which by a law of their own always turn out to be more expensive — never less — than the original estimates.

The circumstances of NASA's past history in the structure of American industry dictated the manner in which the Space Telescope would be built. The first step was to divide the construction of the eventual instrument into components; for instance, the optical part was separated from the support system, and both of these were separated from the scientific instruments. Depending on scientific advice, the various components were assigned specifications, which were then issued to various companies who were invited to submit designs. NASA then awarded contracts to the companies judged to have submitted the best proposals. This procedure has the advantage that it minimizes the cost of maintaining technical expertise within the central organization, but it has the disadvantage of making cost overruns essentially impossible to avoid. Only an organization such as the European Laboratory for Particle Physics (CERN), which has immense technical expertise at the centre, can be expected to reduce cost overruns to less than that provided by the rule of two.

Had the rule been accepted from 1978 onwards, nothing untoward would have happened. But because the Marshall Center was ordered to stay within a budget that was half what it should have been, the project increasingly threatened to develop into what in American parlance is known as a can of worms. There were the usual pitiful signs, familiar surely to everybody who has been involved in the construction of a major instrument. The first step is to rob the budgets of the auxiliary instruments to pay for overruns on the prime mover, as if a ship were robbed of its rudder to pay for the engines. The next step is to push this year's overruns into next year's budget. And the next is to push overruns on the construction into the eventual operating costs. So it went on, along a path that has been followed many times.

Astronomers were given no part in all this. But because of their toehold on the project, through their involvement in the construction of ancillary instruments for the telescope, they could see that

inadequate provision was being made for its eventual operation. In opposition to the Goddard Center, which sought to secure charge of the eventual operation for itself, astronomers preferred the establishment of an independent Space Telescope Science Institute. With strong support from the National Academy of Sciences, the establishment of an independent institute to operate under AURA (Association of Universities for Research in Astronomy) was eventually won. Also won against the opposition of Goddard were AURA's right to appoint its preferred candidate as director of the institute, and the director's right to appoint such manpower as was deemed necessary to provide the software essential for the operations of the telescope. These defeats for bureaucracy carried the lesson that although scientists could easily be dominated so long as they were pitted against each other by that subtle bureaucratic invention, the peer review system, they could not be defeated when they acted in concert, a lesson which particle physicists learned long ago.

By 1983, NASA was in need of a major new project to follow the Space Shuttle, the preferred candidate being the Space Station. There was the problem, however, from NASA's point of view, that many scientists saw the Space Station as being of peripheral importance. To let the Space Telescope continue on its collision course with disaster, a project that astronomers collectively saw as having superb scientific potential, could only lead many of them to speak out frankly against the Space Station. Thus at long last, by 1983, it became necessary for NASA to take the Space Telescope seriously, and to go to Congress with a realistic budget for it. By now, with so much water under the bridge, Congress had little option but to succumb to the buy-in syndrome and so, again at long last, the way was open for the Space Telescope to fulfil its superb potential. I cannot fault anything that happened in the engineering of the telescope from 1983 through to its original projected launch date in October 1986 (which was put back because of the shuttle disaster). If the telescope is the success I expect it to be, it will be in a considerable measure due to those three years of James Beggs's administration of NASA.

So to Act 2. The history of this act will inevitably remain coupled to the manner of the telescope's genesis, with some features of the situation favourable and others not so favourable. It is an advantage that the telescope can be serviced in space. Because of the long lead time associated with such a complex project, exacerbated in this case by the grounding of the Space Shuttle, the scientific instruments will unavoidably be some way behind the state of the art. Servicing in space provides for them to be upgraded

(perhaps they were so improved during the four-year delay in the launch date, but of this I am not aware). Potentially more serious is the low-altitude orbit (about 600 km) into which the telescope had necessarily to be launched to permit rendezvous with astronauts from the shuttle. This brings the instruments into a region of particle bombardment over the South Atlantic. Provision has been made for shielding from these repeated episodes, which one hopes will not seriously reduce the telescope's useful lifetime.

Although I cannot envisage anything very deep for science — to compare, for example, with the group theoretic structures of particle physics — emerging from the planetary research to be carried out with the telescope, such research is certainly to be applauded for its historical associations. Doubtless too, this will be one of the telescope's activities best appreciated by the public. Some play has been made, also from the point of view of public interest, with the possibility of detecting planets moving around stars at distances up to, say, 20 parsecs. A planet of the size of Jupiter moving around a star like the Sun would just about be detectable at this distance, if it could be separated from the diffraction halo of the star. Separation by a subtraction method, however, will require the diffraction halo to be exceedingly steady, a condition that probably will be met so far as the usual seeing problems are concerned. More significant, probably, will be fluctuations in the halo due to slight imperfections in the figuring of the primary mirror.

Other comparatively 'nearby' applications of the telescope could be to the astrometric work on which the establishment of the extragalactic distance scale ultimately depends. The scale begins with the distances of nearby stars, which are essential to calibrating the period-luminosity relation of the Cepheid variables. Once calibrated, the Cepheid variables for a detection limit of +28 should be measureable out to a distance of about 30 megaparsecs. This may be far enough to permit a determination of the Hubble constant (the scale of the Universe) free from the vexed question of local irregularities.

Many projects have been proposed for investigating the Universe at very great distances. Those I have seen are mostly conditioned by the views that cosmologists happen to hold in the year 1990. Should these views prove to be generally

correct, the projects in question will assume historic importance. But if present ideas turn out to be as much in need of modification as those of the past have been, a more general exploitation of the telescope's potential at great distances will be needed. As a broad statement in this regard, I would say the telescope might be expected to perform on objects at a redshift of unity comparably to what telescopes have hitherto revealed at a redshift of one-tenth. Which is to say that, at big distances, an order of

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Work on the primary mirror at Perkin-Elmer's Optical Facility, and (inset) the eponymous Edwin Hubble (NASA & Science Photo Library/Carnegie Institution).

magnitude in precision has been gained.

In their investigations of the isotropy and homogeneity of the Universe, cosmologists have had two considerable surprises. The cosmic microwave background is more uniform than had been expected, whereas the distribution of visible matter is more irregular. The latter is spatial. Little is known concerning uniformity in time. It fits current concepts to suppose that galaxies are uniformly of nearly the same age, but this could very well be an issue on which we are due for a surprise. Technically, the problem is tricky. A galaxy with stars of age 3×10^9 years would be young from a cosmological point of view. Distinguishing stars of age 3×10^9 years by some colour criterion from stars of age, say, 12×10^9 years in 'old' galaxies would be difficult. Yet it might turn out to fit in with proposals for determining variations of chemical compositions among galaxies.

It has been natural for astronomers to make proposals for using the Hubble Telescope on the basis of what they are doing already. If one looks through the list of proposals this is mostly what they are — and of course the observing programme

for the telescope must begin in this way. What one hopes for, however, is that worn tracks, or even partially worn tracks, will eventually be abandoned for what is new and unexpected, which by its nature is at present impossible to discuss. This hope cannot be better expressed than it was in 1946 by Lyman Spitzer:

It should be emphasized, however, that the chief contribution of such a radically new and more powerful instrument would be, not to supplement our present ideas of the universe we live in, but rather to uncover new phenomena not yet imagined, and perhaps modify profoundly our basic concepts of space and time.

Let us come now to the third act, the more distant future. NASA is already proposing a 10m telescope to be sited on the Moon, a proposal to which I give an enthusiastic thumbs down — in part because it would not remotely be a project to which the rule of two could be applied and in part because of the need to explore other possibilities. The \$2,000 million so far spent on the Space Telescope would have been sufficient to construct some 30 or more 8m telescopes on the ground. Ground-based telescopes have the advantages that they can be larger than space telescopes, that they can be serviced continually and that their instrumentation can be always at the frontiers of the possible. Because of the necessarily long lead times, telescopes in space can never enjoy the third advantage. Against all this it must of course be said again that ground-based instruments suffer from seeing difficulties. But this problem will almost surely be overcome in the future. The development of a flexible mirror, adjusted continuously by a kind of Hartmann test, has already reached the stage of essentially conquering seeing problems in the near infrared. It is being confidently said that the technique will be extended to give 0.1 arc-second resolution through the atmosphere on an 8m mirror by the turn of the century — add a few years and say by 2005. So by the projected end of the Hubble Telescope, there are likely to be powerful, less expensive options.

But nothing in these arguments can take away the window of opportunity for the Hubble Space Telescope, a window stretching forward for 15 years. Provided the launch has gone smoothly, and provided there does not turn out to be a problem with pollutants occasioned by the telescope being in a low orbit, the curtain at the end of Act 2 is likely to ring down on a brilliant success. □

Sir Fred Hoyle, c/o St John's College, Cambridge CB2 1TP, UK, was formerly Plumian Professor of Astronomy in the University of Cambridge.

ACKNOWLEDGEMENT. In writing this article I have been indebted — as will scientists and science administrators everywhere — to Robert W. Smith's book *The Space Telescope: A Study of NASA, Science, Technology, and Politics* (Cambridge University Press, 1990).

Who wants a big gun, and why?

The discovery of components for a huge gun apparently being assembled in Iraq has caused a sensation, and has stimulated endless speculation about its nature. Here is some more.

COULD there be a connection between the discovery, at London airport three weeks ago, of capacitors designed for triggering nuclear explosions and the way in which components of a huge gun bound for Iraq have come to light at British and Greek ports in the past few days? Could the triggering devices have been intended not for use with nuclear explosives but, rather, to detonate explosive charges at intermediate points along an overlong gun barrel? What follows is almost entirely speculation, but may go some way to illuminate some of the melodrama of the past few weeks.

The circumstances are these. Ten days ago, British Customs and Excise impounded four massive steel tubes at a British port, announcing that they were sections of a gun barrel. It has since come to light that the manufacturers, in Sheffield, had made a total of 16 tubes of the same design, most of which had already been shipped to Iraq. Initial scepticism of the customs' claim has been exorcised by the discovery last weekend, at the Greek port of Patras, of what seems to be part of the breech mechanism of a huge gun, the barrel of which is reckoned to be more than 160 metres long. Curiously, the device was bound for Iraq on the platform of an open truck.

Those who know about artillery design say that such a barrel is inexplicable on conventional grounds. The most effective high-velocity weapons now in service, notably the South African long-range gun, are built with barrels some 45 times longer than the diameter of the projectiles they are designed to fire. Previous generations of weapons have worked with smaller ratios of barrel-length to calibre — 40 or even less.

That there should be some such limit is understandable. The function of a gun barrel is to allow the continued acceleration of the projectile by the shock-wave from the propellant charge and eventually from simple hydrostatic pressure, which must diminish along the length of the barrel. Plainly there will be a point at which friction between projectile and barrel will be greater than the decreased propellant pressure, at which point the barrel might usefully be terminated.

Moreover, there are also limits to what can be done by increasing the propellant charge; the breech mechanism has to be more robust, but the propagation of a detonation in the solid propellant may be so much slower than that of shock-waves along the barrel that extra propellant

simply goes to waste.

Apparently, the optimum ratio of barrel length to calibre is determined empirically to be roughly constant, yielding a maximum muzzle velocity and thus range, in practice roughly 600 m s⁻¹ and 40 km respectively. (The use of undersized projectiles allows the terminal velocity to be increased, but uneconomically.) The ideal characteristics of the propellant naturally vary with the mass of the projectile. But why should Iraq be bent on building a gun with an over-long barrel? Even with an internal diameter of 1 metre, a barrel 40 m long should give the longest range.

One explanation is that Iraq planned to build a gun in which the muzzle velocity of the projectile would be increased by replenishing the supply of propellant gases at intervals along the length of the barrel. Apparently this has been tried before in the German project, during the Second World War, for building a gun capable of reaching London from the coast of the mainland. Experience seems to have shown that one firing in three caused the gun barrel to explode. It is easily appreciated that the failure of that project may have sprung from the difficulty of timing the explosion of the intermediate propellant charges with sufficient accuracy. The amount of the intermediate charge is important, but the timing of its detonation immediately on the heels of the passing projectile must be crucial.

That is where the detonating capacitors impounded at London Airport earlier in the month may be relevant. It is neither here nor there that the real goods had been replaced by fakes by collaboration between the US and British intelligence agencies — the originals must have been intended for the accurate timing of explosions of some kind. In the Iraqi experiment, if that is what it was, the ideal would be that each section of the gun barrel would be joined to its successor by a short section carrying a cylindrical explosive charge, which would allow of 15 intermediate explosions. It will be interesting to see whether manufacturers in the West have been making to the order of Iraqi agencies cylindrical steel sections of 1 m internal diameter with concentric chambers for carrying internal propellant charges.

Would such an arrangement suffice to give the Iraqi gun a range significant in the Middle East? Because the range of a gun (neglecting atmospheric resistance) is roughly proportional to the square of the muzzle velocity, the auguries are good,

but increasing the range from 40 km to, say, 400 km would require that the muzzle velocity should be increased more than threefold, which is a tall order. Would that be feasible?

Given the complexity of what happens in a gun-barrel, there seems little future in attempts at precise calculation, but a little handwaving may not be inappropriate. The sections of steel tubing impounded by the British authorities all have the same wall diameter (more than 20 cm). Assuming, conservatively, that half the acceleration of the projectile is normally brought about in the first quarter of the barrel, any means of renewing the initial shock wave one quarter of the way along would add to the terminal velocity. And so on, at repeated intervals along the length.

But not *pro rata*. At successive intermediate intervals along the barrel, the relative velocity of the shock-waves from intermediate propellant charges would be reduced, and so would be the efficiency with which momentum is transferred from the gases to the projectile. Again there will be a limiting muzzle velocity — the efficiency of the process will decline to zero when the shock wave cannot catch the accelerated projectile. That limit, a function of the density of gases in the barrel and of their composition, could easily allow the terminal velocity of 600 m s⁻¹ to be exceeded four-fold after 15 intermediate renewals of the shock-wave.

Whether that was really what Iraq was about will probably remain publicly unknown. The originator of the project seems to have been a Dr Gerald Bull, of Canadian origin but with US citizenship, who was murdered by gunmen outside his Brussels apartment three weeks ago. Bull's company, International Space Systems, seems to have placed contracts for the gun components with their British manufacturers. Previously, he had been the driving force for the joint US-Canadian 'HARP' project for sending rocket-assisted projectiles high into the atmosphere and even into orbit and for the development of the South African artillery piece.

Bull's competitors in the arms trade have a great respect for his ingenuity; one said on the telephone this week that "if Gerry said it could be done, I'd believe him and not the others". But the general opinion seems to be that Bull was more interested in gun-launched rockets and in undersized projectiles than in more complicated devices. On this occasion, time may not tell.

John Maddox

Mapping 'frozen accidents'

Gabriel A. Dover

IN 1985 'DNA fingerprinting' arrived¹. Now, from the same laboratory, comes the unveiling of 'fingerprints within fingerprints'². This is a major technical advance in the mapping of fine-grained sequence variation at a minisatellite DNA locus. It opens up the prospect of tracing both paternal and maternal lineages during human evolution, because DNA sequences reflect past mutational events (or 'frozen accidents' as Francis Crick has called them). The achievement ranks with the successful rooting of maternal lineages back to the so-called mitochondrial DNA 'Eve' (ref. 3).

Jeffreys and company² have developed a sophisticated method for isolating and then mapping mutant distributions within single molecules of a given minisatellite DNA locus derived from blood cells and sperm. A minisatellite locus consists of a head-to-tail array of a repeated sequence of DNA in which differences between alleles arise from fluctuations in the number of repeats, due to mechanisms of genomic turnover known as unequal crossingover and/or slippage^{1,4}. DNA fingerprints are unique to an individual and reflect the existence of many scattered loci of minisatellite arrays, all differing in repeat copy-number¹.

The isolation of single alleles of a given array has been achieved with the use of two rounds of PCR (polymerase chain reaction) DNA amplification: the first for the isolation of the original two parental alleles of an individual, and the second for the isolation of much shorter mutant alleles produced by internal deletions of some repeats. The mapping procedure is based on an 'end-labelling' technique to locate the precise positions of any repeats in an array that contained an *HaeIII* restriction site with reference to a *HinfI* restriction site that exists in all repeats in the locus in question (see figure). By this means an exact map of the distribution of *HaeIII* containing repeats within an array (the 'fingerprints within fingerprints') could be determined.

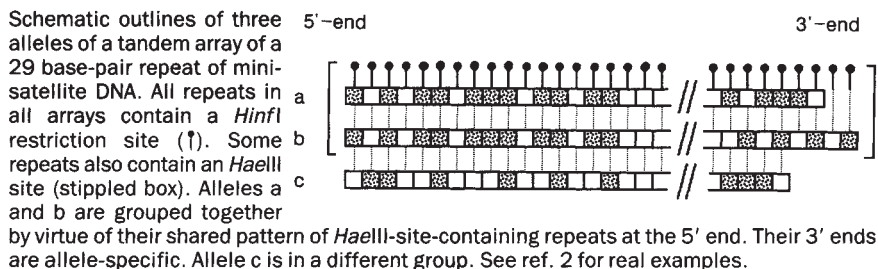
The first application of the technique was the internal mapping of discrete amplified alleles of a locus-specific array of 29 base pair (bp) repeats up to 6 kilobases (kb) long, taken from 80 unrelated individuals of English origin, and from individuals of Amish, French, Mormon and Venezuelan origins. From this survey three features emerged. First, there is sufficient variation in *HaeIII* site-distribution maps for the system to be capable of distinguishing 10²⁰ different allelic states and of measuring somatic and germ-line mutation rates as they occur within a single individual. Second, the alleles can

be grouped into five assemblages according to their *HaeIII* site-dispersion patterns, in particular by the clearly identifiable patterns at the 5' ends of the arrays. The more variable 3' ends are allele-specific (see figure). Third, no assemblage appears to be specific for a given geographical origin of alleles.

What is the basis of the different *HaeIII* site distributions? Previous direct sequencing of different loci of minisatellite arrays showed that some were highly variable in repeat copy-number, but relatively homogeneous in sequence (except for rare mutations being shared by a subset of repeats); other arrays were invariant in

of one progenitor allele was fused to the 3' end of the other allele, as would be expected for a product of unequal crossingover between alleles (non-sister chromatids) by meiotic or mitotic recombination. Interestingly, there is an excess of deletion endpoints within the last 20–30 repeats of the 3' end: the end that cannot be categorized into the haplotype assemblages.

Jeffreys and associates correctly point out that their results cast doubt on the idea that minisatellite repeated sequences are 'hot-spots' for meiotic recombination because of their Chi-like generalized recombination signals (see refs 1 and 4 for previous discussions of Chi). Arrays may be evolving along haploid lineages in splendid isolation, with some arrays accidentally wandering off into oblivion, whereas others are the lucky survivors from which all extant arrays are related by



copy-number but variable in sequence with unique mutations in given repeats¹. These contrasting situations can be explained by models of repeat sequence evolution in which variant repeats diffuse through an array by replacement of non-variant repeats, as the inevitable consequence of mechanisms of stochastic gain-and-loss of repeats such as unequal crossingover and slippage^{3,6}. Hence, the partial spread of *HaeIII* sites in an array would reflect the incomplete homogenization of a relatively recent mutation generating the *HaeIII* site, with particular distribution patterns reflecting the localized activities of unequal crossingover in different parts of an array.

The next question is why the alleles of unrelated individuals fall into one of five assemblages. The answer depends on whether unequal crossingover is constrained to occur between mitotic sister chromatids, or even within single chromatids. If this is the case then similarities in *HaeIII* distribution maps would reveal the descendants of a particular haploid chromosome lineage (haplotype) and thereby afford the opportunity to monitor human origins and dispersal.

To test this possibility, Jeffreys and colleagues ingeniously amplified single molecules of sperm and blood cell deletion-mutants (an achievement in itself), after suitable size fractionation of the DNA of one male. In 64 germ-line deletion-mutants and 42 somatic deletion-mutants there was no instance in which the 5' end

descent. To explain both the complete homogeneity of *HinfI* sites in all arrays and the partial diffusion of *HaeIII* sites in each array, the Ur (ancestral) chromosome would have had to have carried both sites, with unequal crossingover (or other mechanism), restricted now to single or sister chromatids, diffusing the sites through an array. Although the repeat unit is present in some species of higher ape, there are only two copies of it, both of which lack the *HaeIII* site (Jeffreys, personal communication). So it would seem that the mutation generating the *HaeIII* site, and the Ur chromosome with its amplified array, arose after the hominid-great ape divergence. (If all chromosomes belonging to a given assemblage (haplotype) are descendants of one progenitor, then their different 3' ends need explaining. This could be due to a hot-spot of activity of the mutational process continually reshuffling the patterns of variant repeat units at that end.)

But is this the only true reading of events? If some of the sharing of the *HaeIII* site distribution patterns between alleles in a given assemblage were the result of inter-allelic micro-gene conversion (sometimes also called 'segmental' gene conversion) which had been restricted to the 5' ends, then chromosome lineages would not necessarily be evolving independently. Localized micro-gene conversions, swapping patches of sequence only and creating mosaic genes, are well documented in human globin, immuno-

globulin and HLA genes (for example, see ref. 7). Such mosaic alleles, not having a change in length, would not have been isolated in a size fractionation procedure that screened for much smaller deletion products of approximately 0.5–2.5 kb. The same would be true for the products of any inter-allelic unequal crossingover that involves a crossover displacement of only a few of the 29 bp repeats at a time.

With either of these mechanisms of sequence 'crosstalk', chromosomes are effectively evolving as a cohesive unit, roped together in a slow boat to nowhere. Significantly, 'crosstalk' is essential to explain patterns of mutant distribution amongst the five chromosomally separated arrays of human rDNA, in addition to the lottery of chromosome loss^{8,9}. Only occasional sequence transfers need take

place between chromosomes to significantly increase their genetic identity. Whatever the relative roles of the two models — lucky chromosomes or lucky sequences — the past is still within us. □

Gabriel A. Dover is in the Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK.

1. Jeffreys, A.J., Wilson, V. & Thein, S.V. *Nature* **314**, 67–73 (1985).
2. Jeffreys, A.J., Neumann, R. & Wilson, V. *Cell* **60**, 473–485 (1990).
3. Cann, R.L., Stoneking, M. & Wilson, A.C. *Nature* **325**, 31–36 (1986).
4. Dover, G.A. *Nature* **342**, 347–348 (1989).
5. Smith, G.P. *Science* **191**, 528–535 (1976).
6. Dover, G.A. *Nature* **299**, 111–117 (1982).
7. Howard, J.C. *Nature* **332**, 588–590 (1988).
8. Arnheim, N. in *Evolution of Genes and Proteins* (eds M. Nei & R.K. Koehn) 38–61 (Sinauer, Sunderland, Massachusetts, 1983).
9. Dover, G.A. *Genetics* **122**, 245–252 (1989).

ASTRONOMY

Meteorite parent bodies

Clark R. Chapman

CERTAIN objects in interplanetary space have been studied for many decades by two different groups of scientists. Like the blind men describing an elephant, their perspectives are so distinct that one group (planetary astronomers) has long called them 'minor planets' or 'asteroids' whereas the other (meteoriticists and cosmochemists) calls them 'meteorite parent bodies'. Some intriguing inferences about the parent body of a specific meteorite now come from the meteoriticist section of the community (K. Keil *et al. Geochimica et Cosmochimica Acta* **53**, 3291–3307; 1989).

The vantage points of the two groups could not be more dissimilar. Astronomers, who cannot even resolve the disks of asteroids, study their global and disk-averaged properties. Meteoriticists analyse hand samples using laboratory techniques, generally on the microscopic level. And whereas astronomers observe specific charted, named bodies, meteoriticists — who cannot specify particular asteroidal sources — interpret their results in terms of hypothetical parent bodies. In the past decade, it has become clear that these two classes of bodies are the same — almost all meteorites must be derived from asteroids, although specifically which asteroids remains unknown.

Late in 1991, the Galileo spacecraft will fly past the small asteroid Gaspra, finally bridging the enormous gap in resolution between the telescopic and the microscopic. Meanwhile, however, meteoriticists have been making ever more specific inferences about the traits of the asteroidal bodies from which specific meteorites were derived and about their evolutionary histories.

Keil and colleagues have studied an

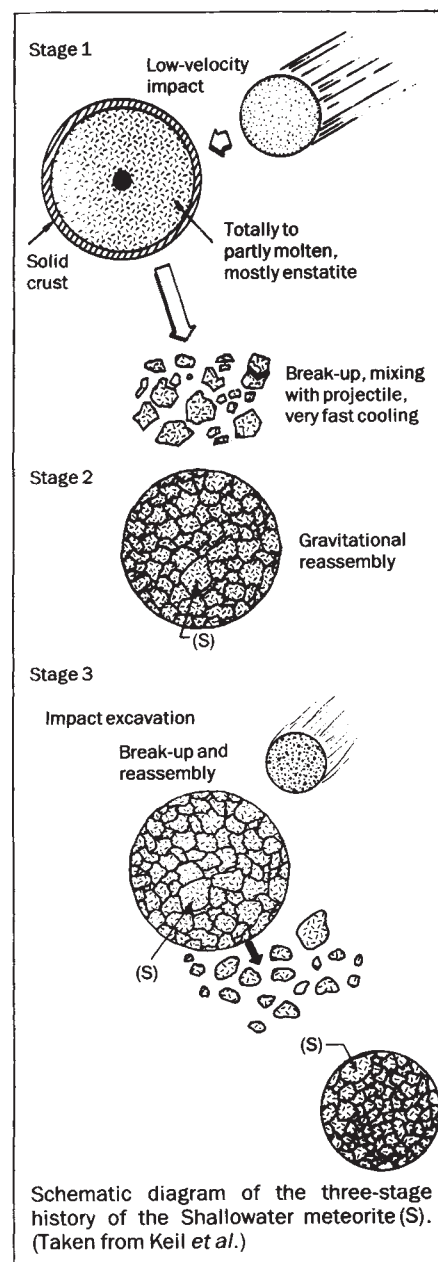
unusual meteorite, Shallowater. It is one of a small number of 'aubrites', which consist mainly of the mineral enstatite, a magnesium-rich pyroxene. Shallowater also consists of bits of other minerals like olivine and plagioclase, plus a few per cent of metal grains.

Unlike many meteorites, which are moderately pristine remnants of materials of the solar nebula, the aubrites are differentiated meteorites — their composition reflects extensive igneous processing and separation of components due to melting of the parent asteroidal body. An over-simplified picture of a differentiated body is one in which metallic material sinks to form a core and the lightest magmas float to form the body's crust.

Keil *et al.* have analysed the compositions and structures of the crystal grains that make up the Shallowater meteorite. They conclude that the originally molten material underwent a complex, three-stage cooling history. The first stage lasted only hours, during which the material cooled by several hundred degrees from a starting value above the melting temperature of enstatite, which is 1,580 °C. Evidence for this very rapid cooling comes from the existence of microstructures within intergrown crystals of different compositions.

In the second stage, Shallowater cooled very slowly. It took millions of years to cool from 712 °C to 680 °C, as indicated by the profiles of nickel concentration within the metal grains. Meteoriticists frequently determine the cooling rates of meteorite parent bodies from such an analysis of crystal patterns within the metallic component of meteorites.

Finally, there was another period of very rapid cooling that 'froze in' the nickel



concentrations used to infer the very slow cooling. In particular, the temperature seems to have dropped from 680 °C to only 300 °C in just two years.

The problem that Keil and co-workers believe they have solved is this: how was Shallowater made, with its diversity of minerals, and processed through such a complex cooling history? Many *ad hoc* schemes might work but, for a mechanism to be plausible, it should make physical and chemical sense in the context of what is known about interplanetary space. The process should also generate a large quantity of the material, so that there is a reasonable probability that it would be found among the rather limited sampling of meteorites in our collections.

Keil *et al.* believe that they have identified such a scheme. They suggest that the meteorite was actually made from material originally in two separate asteroids of somewhat dissimilar composi-

tion. They argue that a small cold, solid asteroid collided with a larger, molten asteroid (about 100 km in diameter). The collision resulted in the break-up of both bodies and enabled mixing of the cold and molten materials to take place. The fragments ejected into space, however, lacked sufficient velocity to escape into independent orbits around the Sun. Rather, after only a matter of hours, the debris fell back together again, forming a 'rubble-pile' asteroid. The very rapid cooling would have taken place mainly during the hours when the fragments were scattered in space. Then the Shallowater meteorite was buried deep within the rubble pile. If the rubble pile was about 100 km in diameter, then, for cooling to have nearly stopped for millions of years, the Shallowater precursor would have had to be buried to a depth of about 40 km. The third stage of rapid cooling requires another collision in order to bring the Shallowater material near to the surface; once there, it would cool quite rapidly to ambient temperatures.

Is their proposal an improbable fantasy, or does it make sense when looked at from the asteroid-astronomy side of the house? The qualitative idea that collisions often form rubble-pile asteroids has become increasingly in vogue since it was suggested more than a dozen years ago. For instance, Housen and Holsapple have devised a quantitative model for the collisional formation of rubble piles (*Icarus* **84**, 226–253; 1990).

Less certain is the idea that both solid and molten bodies existed in the asteroid belt at the same early epoch of the formation of the Solar System. There must have been many of both, and they must have lasted for sufficient time to yield high enough collision probabilities to make two-component (melt/xenolithic) mixtures that have a chance of winding up in our museums on Earth. Keil, however, is not the first to explain a type of meteorite by invoking two-component mixtures of two disparate asteroids. In 1985, John Wasson and Alan Rubin published such a model for the mesosiderites, a form of stony-iron meteorite (*Nature* **318**, 168–170; 1985).

In both cases, the chances of actually mixing the different material would be enhanced if collisions occurred at lower velocities than the now typical 5 km s^{-1} . (High-velocity impacts simply vaporize and eject almost all of the projectile material.) Perhaps such models will help constrain our knowledge of the thermal and orbital histories of planetesimals early in the history of the Solar System, so that we may then better understand how the terrestrial planets formed. □

Clark R. Chapman is in the Planetary Science Institute, 2421 East 6th Street, Tucson, Arizona 85719, USA.

Evolution and fisheries

William J. Sutherland

It is considered good fisheries practice to harvest larger individuals and conserve the smaller ones; indeed, there is legislation in many countries to ensure that stocks of shellfish, whales and fish are protected in just this way. But Richard Law^{1–3} and his colleagues have shown how such harvesting can exert an evolutionary selection pressure that may result in genetic changes in life history: alterations in life-history variables such as growth rate or age of maturity can reduce long-term yields from fisheries.

Populations may be overexploited by recruitment overfishing or by growth overfishing⁴. In recruitment overfishing, a high mortality in all age classes ensures that few fish reach maturity and subsequent recruitment will be low, a process that was probably responsible for the collapse of the herring industry in Europe in the mid-1970s. In growth overfishing, many individuals are caught when young, despite the fact that leaving these fish to mature would result in a higher biomass; this process is the more important for most fish populations in the North Sea. The recommended solution to both types of overfishing is the same: exploit the larger fish and conserve the smaller ones.

The life history of a species has been shaped by selection pressures. Fishing mortality could act as a new selection pressure producing genetic change⁵: if, for example, larger individuals are more vulnerable to harvesting, then early maturity and slower growth might be favoured. An example of strong selection pressure imposed by fishing is found in North Sea cod, haddock and plaice, for which the fishing mortality exceeds natural mortality once the fish have been recruited into

the fishery. This selection may result in genetic change if life history characters are heritable. The evidence shows that traits such as growth rate and age of first breeding are heritable and that cultivated populations of shellfish and fish respond to selection for these characters⁶.

Law and Grey¹ propose that changes in the pattern of fishing mortality on the Arcto-Norwegian stock of cod have indeed resulted in changes in life history. For at least one thousand years this stock has been fished on the spawning grounds near the Lofoten Islands. Ocean currents disperse the eggs to the Barents Sea and the seas around Spitsbergen, where the young cod remain and feed until they reach sexual maturity; thereafter they undertake annual migrations to the spawning grounds. This stock matures unusually late — at about eight years old — which is as might be expected when high mortality is associated with breeding. Since the construction of deep-sea trawlers in the 1930s, these fish have been exploited on their feeding grounds. The probability of a fish surviving the period between three and eight years of age used to be 40 per cent, but as a result of this new intensive fishery on all age classes, this probability is now reduced to 2 per cent. Such a shift in fishing mortality must affect the selection pressures acting on the stock, and in accordance with this there has been a gradual shift towards earlier maturation, with more young fish entering the breeding stock. Changes in life history also emerged from a study of five species of Pacific salmon in British Columbia⁷.

In a rigorous attempt to show whether harvesting can result in changes in life histories and yield, selection experiments



Patrick Sutherland

Checking salmon butts — will current harvesting practices affect long-term yields?

have been set up in the laboratory using populations of the cladoceran *Daphnia magna*³. Every four days the populations were culled by passage through a sieve; in some populations many of the large individuals were removed, whereas in others the small ones were selectively harvested. About 40 per cent of the individuals within the selected size classes were removed with each harvest. After carrying out this procedure for 152 days (about ten generations), the experiment continued for a further 68 days during which the *Daphnia* were culled independently of their size, to see whether any changes persisted in the absence of selection. The yield from the populations in which the larger individuals were harvested declined markedly over time, and this low yield persisted throughout the size-independent harvesting. Harvesting the smaller individuals did not affect yields. Rearing newborn individuals in isolation showed that these patterns were associated with genetic changes in life history. The *Daphnia* from the population in which small individuals had been culled initially grew rapidly and reproduced when large, but those populations from which large individuals were harvested grew slowly and reproduced at a smaller size. So these experiments confirm that the processes attributed to the cod¹ and salmon⁷ populations really can occur in harvested populations.

Fisheries scientists use population dynamics to determine strategies for producing the highest yields or the greatest financial returns. But if life histories are altered as a result of harvesting, then there is a risk that their policies may bring about a long-term decline in yields. A simplified model, which included evolutionary change, was used by Law and Grey¹ to calculate the consequences of managing the Arcto-Norwegian stock of cod in different ways and to show how long-term yield is maximized by concentrating on the spawners on the Norwegian coast. For example, assuming an instantaneous fishing mortality of 0.2 in the fishery on the spawning grounds and 0.6 on the young fish in the Barents Sea (which probably resembles present policy), the model predicts that the stock will evolve until individuals mature at four years old and that exploiting the population will give an equilibrium yield of 0.21 million tonnes a year. But by changing the policy so that all the fishing is on the spawning stock and at a level imposing a mortality rate of 0.6, the stock should then evolve until individuals mature at nine years old and the harvest will give an equilibrium yield of 1.44 million tonnes a year.

Size-selective harvesting effecting life-history changes may also occur in natural populations as a result of predation⁸. Oystercatchers, crabs and starfish select medium-sized mussels, ignoring the larger ones that are too difficult to open, yet they

are capable of opening most cockles. The selection brought about by these different patterns of mortality has probably altered the life histories of these two bivalve species: cockles grow slowly and mature early, and mussels grow quickly and delay reproduction.

This research provides another depressing component to fisheries management. Combining fisheries models with economic models shows that it makes economic sense to overexploit populations if many individuals have access to the same fishery⁹. For example, almost all the fish stocks around the British Isles are well below the level at which they would produce the highest yields of fish and the largest profits. It has proved difficult to persuade politicians to take a medium-term view of this problem. These studies

SOLAR OBSERVATIONS

Fine structure on the Sun

Nigel Weiss

AT last, details of motion and magnetic structures on the Sun can now be seen at a resolution of only a few hundred kilometres. The latest results¹, reported on page 842 of this issue, were obtained as soft X-ray images of the solar corona using a rocket-borne X-ray telescope. The images show structure down to a scale of about 0.75 arcsecond (1 arcsec corresponds to 725 km) with indications that self-similar fine structure persists down to much smaller scales. The striking full-disk image, reproduced in false colour on the cover of this issue, shows several active regions and a large flare with fine bright filaments, together with eruptions on the outer edge of the disk and isolated X-ray bright points — this is perhaps the most dramatic image of solar activity seen so far.

The Normal Incidence X-ray Telescope (NIXT), developed jointly by the Smithsonian Astrophysical Observatory and IBM, has multilayer coatings to increase its X-ray reflectivity. It has been flown three times aboard rockets of the National Aeronautics and Space Administration launched from the White Sands Missile Range in New Mexico. The second flight yielded some flare images² but the remarkable new images were all obtained on the third flight, launched on 11 September 1989. The photographs (in an Fe XVI line at a wavelength of

suggest that it may be necessary to take an even longer-term view and to rethink the way in which wild populations are harvested. □

William J. Sutherland is in the School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

1. Law, R. & Grey, D.R. *Evol. Ecol.* **3**, 343–359 (1989).
2. Law, R. & Grey, D.R. *Size-structured Populations* (eds Ebenman, B. & Persson, L.) 140–154 (Springer, Berlin, 1988).
3. Edley, M.T. & Law, R. *Biol. J. Linn. Soc.* **34**, 309–326 (1988).
4. Pitcher, T.J. & Hart, P.J.B. *Fisheries Ecology* (Croom Helm, London, 1982).
5. Blythe, S.P. & Stokes, T.K. *J. Math. appl. Biol. Med.* (in the press).
6. Gjerde, B. *Aquaculture* **57**, 37–55 (1986).
7. Richer, W.E. *Can. J. Fish. aqu. Sci.* **38**, 1636–1656 (1981).
8. Seed, R. & Brown, R.A.J. *Anim. Ecol.* **47**, 283–292 (1978).
9. Clarke, C.W. *Bioeconomic Modelling and Fisheries Management* (Wiley-Interscience, New York, 1935).

63.5 Å) show structures in the solar corona, up to 200,000 km above the surface, at temperatures of 2–3 million K (see figure for example). Bright areas denote active regions and the extremely fine-scale and topological complexity of these magnetically active structures can easily be seen. Enlarged images of the flare region reveal structure on scales running from several arcminutes down to the limits of resolving power of the telescope, with every indication that a self-similar cascade persists down to even smaller scales. This is an enormous improvement on the resolution of the classic X-ray pictures obtained from space during the manned Skylab mission in 1973–4³.



X-ray image of eruptions on the outer edge of the solar disk (from ref. 1: courtesy of L. Golub).

Golub *et al.*¹ analyse the fine structure in the region of the solar flare, drawing attention to the ribbon-like features formed as matter pours down on to the photosphere from above. The ribbons are connected by magnetic loops and energy is released as magnetic field lines reconnect at the top of these loops. An enormous amount of energy (about 10^{25} J) is released and this raises the question: how can the energy residing in a magnetic field extending over huge distances — comparable with the solar radius (nearly 110 times the radius of the Earth) — be released in just a few minutes to supply the energy radiated from a flare? Ohmic dissipation over very small length scales must occur as the field lines reconnect: and that is just what seems to be happening in the X-ray images. New theoretical work is clarifying the way in which such rapid reconnection in thin current sheets around magnetically neutral points occurs to feed the flare outburst¹.

We know from the Einstein space mission in 1979 that, in other stars, X-ray emission is correlated with magnetic activity. There are two ways of heating their coronae. One is by dissipating energy carried outwards by hydromagnetic (Alfvén) waves; the other is by ohmic dissipation in regions with strong currents. Magnetic loops emerging in active regions are shunted to and fro by turbulent convective motions below the solar surface. It is widely believed that when the magnetic fields are sufficiently distorted, they collapse catastrophically to form a new configuration with thin current sheets in which energy can be rapidly dissipated^{2,3}. The X-ray images corroborate this picture.

Other observations at high resolution illuminate the interaction between magnetic fields and convection. The Spacelab-2 mission in 1985 yielded white-light images with exceptional resolution. At first, the results seemed disappointing but the team at Lockheed Palo Alto Research Laboratories realized that they were contaminated by the solar oscillations of period 5 min. When the solar oscillations were filtered out, the small-scale convective motions — the solar granulation — were seen with unprecedented clarity. The same technique has now been applied to observations obtained from the ground at the Pic du Midi and at the Swedish Observatory at La Palma^{4–6}. These new observations show

details of the velocity pattern⁹ and of intense magnetic fields, which are typically confined to small flux tubes that nestle between the individual convective granules. The motions of these flux tubes are believed to be responsible for disturbances which heat the corona.

The behaviour of the Sun cannot be modelled in experiments on Earth, although numerical simulations of com-

pressible convection and its effect on magnetic fields have allowed some progress to be made. Explaining the observations, with ever-increasing evidence of fine structure, remains a challenge to theoreticians. □

Nigel Weiss is in the Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 9EW, UK.

OCEANIC CRUST

New images for old faults

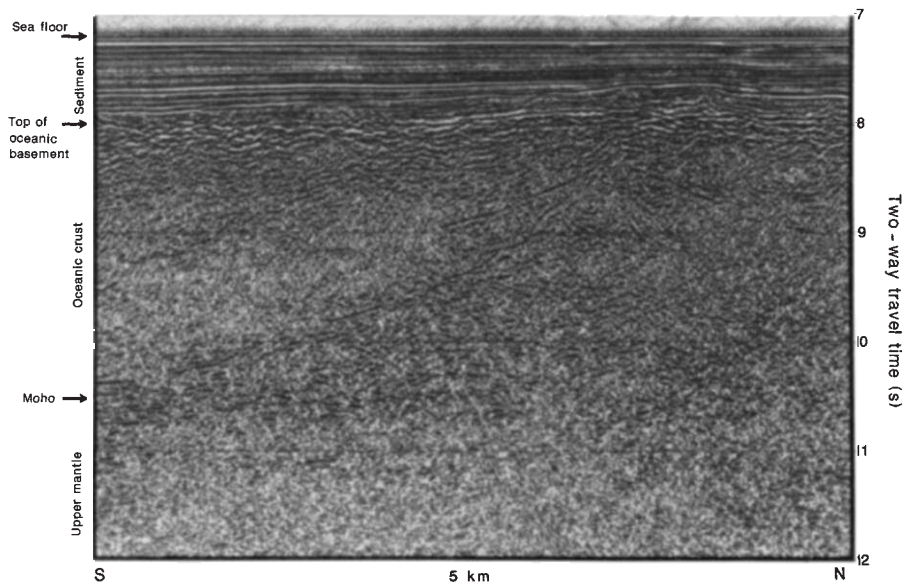
Sean C. Solomon

THE traditional view of the seismic structure of oceanic crust is one of vertical layering produced by the magmatic processes which form the crust at a spreading centre. Two new studies, reported on page 855 of this issue¹ and in the May issue of *Geology*², use multichannel seismic profiling to reveal a variety of non-horizontal structures in old oceanic crust. The interpretation of these structures is changing our view of the formation and subsequent tectonic deformation of oceanic lithosphere.

Multichannel seismic (MCS) profiling is an important geophysical tool in exploration for oil and gas, but has hitherto been used only sparingly in the study of mature oceanic crust. A ship tows a long streamer of closely spaced hydrophones to record signals from a repetitive sound source, usually an array of air guns: by summing the hydrophone traces from suitably grouped records, the phases reflected off structural boundaries within the crust can be enhanced. The travel time taken against position and source range for such phases permits the imaging at depth of the reflecting boundary. MCS profiling has

been used widely in the investigation of continental margins³, and has also been carried out along the axes of some mid-ocean-ridge spreading centres⁴.

Bull and Scrutton¹ describe the imaging with MCS techniques of fault structures within oceanic crust approximately 70 million years old. Their survey area, in the central Indian Ocean basin, is part of a broad region of intraplate compressional deformation manifested in undulations and faulting of the oceanic crust⁵, earthquakes⁶ and geoid anomalies⁷. The MCS images confirm that the faults are high-angle and reverse in nature, with approximately equal numbers dipping to the north and to the south. From the strikes (about 100° E), dip angles (35 – 45°) and lengths (generally less than 10 km) of the imaged faults, Bull and Scrutton infer that the structures are reactivated normal faults originally formed near a spreading centre. Large earthquakes along the present axes of slowly spreading mid-ocean ridges have mechanisms indicating normal faulting on plates dipping at 40 – 55° , and seismic moments and source durations consistent with typical fault lengths of



MCS profile of oceanic crust along a 145-Myr isochron in the western north Atlantic². The vertical axis is two-way travel time. Several prominent southward- and northward-dipping reflectors in the middle to lower crust are visible. (Courtesy of R. S. White.)

- Golub, L. *et al.* *Nature* **344**, 842–844 (1990).
- Martens, P. C. H., Golub, L. & Herant, M. *Proc. Cargèse Workshop on Plasma Phenomena in the Solar Atmosphere* (in the press).
- Stix, M. *The Sun* (Springer, Berlin, 1989).
- Priest, E. R. (ed.) *Basic Plasma Processes on the Sun* (Kluwer, Dordrecht, in the press).
- Moffatt, H. K. & Tsinober, A. (eds) *Topological Fluid Mechanics* (Cambridge University Press, 1990).
- Brandt, P. N. *et al.* *Nature* **335**, 238–239 (1988).
- Rutten, R. J. & Severino, G. (eds) *Solar and Stellar Granulation* (Kluwer, Dordrecht, 1989).
- Stenflo, J. O. (ed.) *Solar Photosphere: Structure, Convection and Magnetic Fields* (Kluwer, Dordrecht, 1990).
- Edmunds, M. G. *Nature* **344**, 385–386 (1990).

about 10 km (ref. 8), supporting their inference.

One of the most interesting results of the survey is that the high-angle faults appear to extend, in an approximately planar geometry, to the base of the crust and possibly into the uppermost mantle. Earthquakes at mid-ocean ridges have centroid depths consistent with faulting of the entire crustal column⁸, but only along ridges spreading at half-rates of 20 mm per year or less. The spreading half-rate at which the crust of the MCS survey was generated, however, was about 80 mm per year⁹. Ridges spreading at such fast rates are characterized by only small-magnitude earthquakes confined to the uppermost few kilometres of the near-axis crust¹⁰. Bull and Scrutton speculate that the deepest imaged portions of the faults in their survey may not have originated at the ridge axis but instead formed by downward extension of upper crustal faults during the recent episode of lithosphere compression. Reverse-faulting earthquakes in the central Indian Ocean basin have centroid depths of 20–40 km (ref. 6), suggesting that slip on these faults nucleates in the strong layer of the upper mantle and that these fault structures may extend much deeper than even the best current images indicate.

In a related MCS profiling study, White *et al.*² report a variety of non-horizontal reflectors within the Mesozoic crust of the western north Atlantic near the Blake Spur fracture zone. In a profile following the direction of spreading, they have imaged a number of planar reflectors. The most prominent are within the lower crust, but some appear to cut through the entire crustal column. Dip angles are 20–40°, with two-thirds of the lower crustal reflectors dipping eastward towards the ridge. Along an orthogonal isochron line, planar low-angle reflectors were also imaged (see figure). Dip angles are 10–30°, with southward dips most common.

Although a single set of obliquely trending structures is a possible interpretation, limited data from profiles at other azimuths in the same region lead White *et al.* to conclude that the profiles in the spreading and isochron directions are generally imaging two different sets of structures. They interpret the structures imaged in the 'spreading direction' profile as normal faults formed near the ridge axis. The dip angles of the reflectors are somewhat less than the dip angles indicated by the mechanisms of ridge axis earthquakes⁸, but the difference could be attributed to a general steepening of the faults in the upper crust, as is seen for some of the reflectors², or to rotation of fault-bounded crustal blocks outwards from the seismically active zone of the median valley. There is, too, the possibility that some of the reflectors, particularly those terminating in the middle to

Still standing after the earthquake



A truck escapes untouched on the lower roadway of the Nimitz freeway in Oakland, California, following the Loma Prieta earthquake of 17 October 1989. In the 1.4-km-long section of the freeway that collapsed in the earthquake, only the segment shown survived; the lower roadway here is supported by three pillars at each end instead of the usual two. On page 853, an analysis of aftershock recordings sheds some light on the failure of the Nimitz freeway; it seems that the presence of three-pillar supports strengthened the segment sufficiently to prevent its collapse.

upper crust, may be relics of ridge-axis magmatic processes².

White *et al.* interpret the planar structures observed along isochron lines as thrust or reverse faults. Lithospheric thermal stress models do predict compressive stress in young oceanic crust¹¹, but near-ridge earthquakes with compressive mechanisms appear to occur on reverse faults¹² with dip angles steeper than the 10–30° seen for planar reflectors along isochrons². This discrepancy may only be apparent, however, as a large fraction of the reverse-faulting earthquakes in young oceanic lithosphere occurs on fault planes oblique to an isochron¹, so that the true dip angles of fault surfaces imaged along isochrons may be significantly larger than those inferred from the MCS profiles.

An unresolved issue is why the imaged structures are such prominent reflectors of seismic energy. A distinct contrast in seismic velocity or density is implied. The presence of fluids in the fault zones has been suggested as an explanation for the structures in the Central Indian Ocean Basin¹; frictional heating and fault-controlled hydrothermal circulation may contribute to localized, anomalously high heat flow in the region⁵. But such an explanation does not obviously apply to

the structures imaged in the crust of the western Atlantic. Perhaps mineralization during an earlier episode of hydrothermal circulation has left a permanent difference in physical properties compared with those of the surrounding crust. Measurements of the physical properties of rocks from fault zones in ophiolites may help to resolve this question. □

Sean C. Solomon is in the Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

1. Bull, J.M. & Scrutton, R.A. *Nature* **344**, 855–858 (1990).
2. White, R.S. *et al.* *Geology* **18**, 462–465 (1990).
3. Watkins, J.S., Montadert, L. & Dickerson, P.W. *Geological and Geophysical Investigations of Continental Margins*, Mem. 29, (Am. Ass. Petrol. Geol., Tulsa, 1979).
4. Detrick, R.S. *et al.* *Nature* **326**, 35–41 (1987).
5. Weissel, J.K., Anderson, R.N. & Geller, C.A. *Nature* **287**, 284–291 (1980).
6. Bergman, E.A. & Solomon, S.C. *Phys. Earth planet. Inter.* **40**, 1–23 (1985).
7. McAdoo, D.C. & Sandwell, D.T. *J. geophys. Res.* **90**, 8563–8569 (1982).
8. Huang, P.Y. & Solomon, S.C. *J. geophys. Res.* **93**, 13445–13477 (1988).
9. Sclater, J.G. & Fisher, R.L. *Geol. Soc. Am. Bull.* **85**, 683–702 (1974).
10. Riedesel, M., Orcutt, J.A., Macdonald, K.C. & McClain, J.S. *J. geophys. Res.* **87**, 10613–10623 (1982).
11. Bratt, S.R., Bergman, E.A. & Solomon, S.C. *J. geophys. Res.* **90**, 10249–10260 (1985).
12. Bergman, E.A. & Solomon, S.C. *J. geophys. Res.* **89**, 11415–11441 (1984).

Fresh pathways to follow

Dani P. Bolognesi

THE intricate mechanisms used by the immune system to recognize antigens are of great interest not only to immunologists but also to those involved in vaccine development. The trend over the past several years has been to group antigens into two categories according to the response they induce in the immune system (for review, see ref. 1). Non-replicating entities that enter the antigen-presenting cell from outside (exogenously) are processed in the endosomal compartment and presented at the cell surface in association with class II major histocompatibility complex (MHC) molecules. This activates CD4⁺ helper T cells which, among other things, are required for the production of antibodies as well as CD4⁺ class II-restricted cytotoxic T lymphocytes (CTL). Alternatively, antigens enter a different processing pathway which is functional for proteins such as viral proteins that are synthesized inside the cell (endogenously). In this case, association is with MHC class I molecules and it is this complex that primes CD8⁺ CTL. From the standpoint of immunization with anything other than an infectious agent, it has been something of a dilemma to design non-replicating immunogens that could allow processing and presentation by both class I and class II MHC. But several studies have indicated that these pathways are not as clear-cut as was once thought. For example on page 873 of this issue², Takahashi and colleagues now report that a unique subunit immunogen which induces neutralizing antibodies against human immunodeficiency virus (HIV) (B. Morein, personal communication) can also prime MHC class I-restricted HIV-specific CD8⁺ CTL.

Other examples are to be found in the generation of MHC class I-restricted CTL against soluble ovalbumin³ and against influenza virus peptides attached to 'lipid feet'⁴. On the other side of the coin is the novel observation that influenza virus proteins processed by the endogenous pathway can associate not only with MHC class I but also with class II molecules⁵. Association of endogenously synthesized antigens with class II MHC also occurs in the case of hepatitis B virus⁶ and for HIV envelope antigens synthesized inside the cell⁷. In both of these examples, processing occurs in the endosomal compartment and is therefore distinct from the influenza case⁵. As investigation continues along these lines, models for antigen processing and presentation will probably converge further (see box), as already anticipated^{1,7,8}.

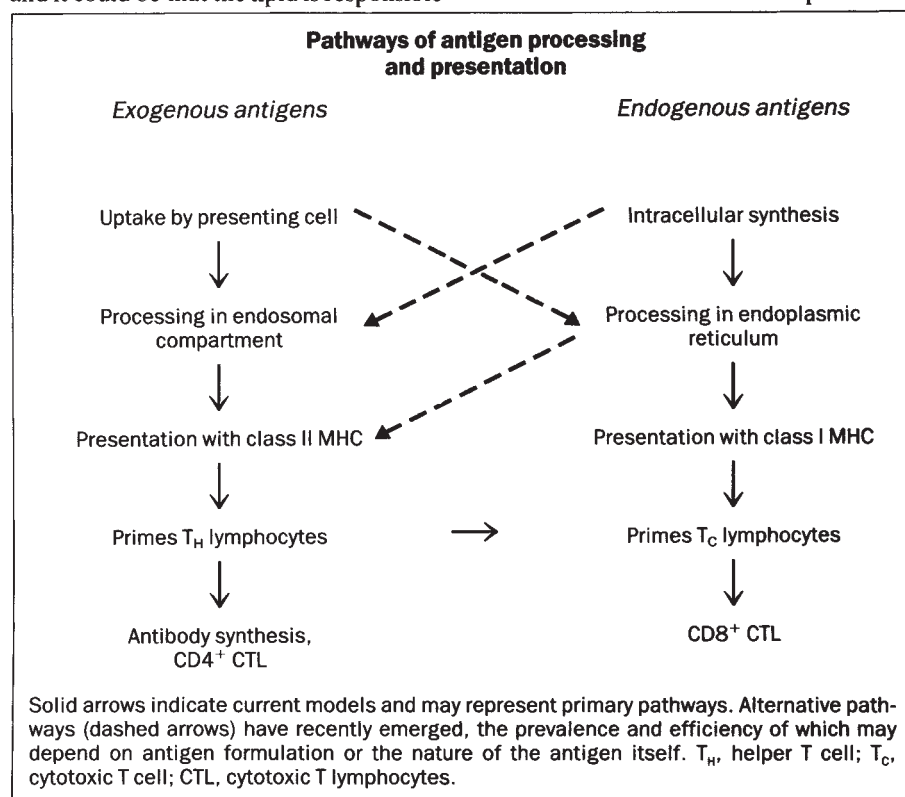
These studies emphasize the need to understand better the different ways in

which exogenous and endogenous antigens are processed and compartmentalized in cells. With regard to exogenous antigens, one might look to some mammalian viruses for clues. Viruses with membrane envelopes can penetrate cells by two mechanisms, one by fusion within endocytic vesicles, the other by direct fusion with the plasma cell membrane. It would be of interest to know how the appropriate entry route is selected. The answer may lie in the viral fusogenic proteins themselves and a more satisfactory definition of their interactions with the cell surface may help us to make cells respond selectively even to inert subunit immunogens. Manipulations with inactivated influenza virus preparations already look promising⁹. Influenza virus normally infects by fusion at low pH in endocytic vesicles, but heat inactivation of the virus causes fusion that allows the viral antigens to bypass the endosomal compartment and so to induce MHC class I-restricted CTL. Curiously, an opposite effect is also possible: for example, live measles virus fuses directly with the plasma membrane and yet preferentially induces a class II-restricted CTL response¹⁰.

The precise mechanisms by which the influenza peptide derivatives⁴ and the immune-stimulating complexes (ISCOMs) of Takahashi *et al.*² enter cells are not known. Both are associated with lipid, and it could be that the lipid is responsible

not only for entry, but also for the susceptibility to processing by the endogenous pathway. On the other hand, as recognition by class II-restricted CTL of endogenously synthesized HIV envelope glycoproteins needs a membrane anchor⁸, lipid/protein interactions must be important for delivery of endogenous antigens to the compartments where processing and association with MHC class II products can occur. One could speculate that antigens that remain associated with membranes (both within and outside the cell) may be destined for the class II pathway, whereas those that do not (but can still penetrate cells in the case of exogenous antigens) are channelled to the class I pathway. This might be the case for viral core versus certain envelope antigens, for example. It could also be relevant to measles virus if the dominant CTL epitopes reside in components that associate with membranes.

For HIV vaccine researchers this is welcome news, principally because of the widely held belief that protective immune responses must be directed against both free virus and infected cells. Until now, only attenuated HIV preparations or replicating recombinant vectors were candidates, but the former are excluded because of safety considerations and the latter are still in the developmental stages. The prospect that subunit or peptide immunogens could be endowed with properties enabling them to enter either or both of the main processing pathways and stimulate a full range of immunity will receive considerable attention. One wonders too whether the protective



response obtained with recent inactivated simian and human immunodeficiency viral vaccines¹¹⁻¹³ could have included a CTL component, because the replicating forms of these viruses fuse directly with the plasma cell membrane¹⁴. But the use of killed HIV preparations for human vaccines carries considerable risk, and structures like the ISCOMs bearing only selected viral components may be attractive alternatives. In this regard the success of using ISCOMs to introduce whole viral proteins into cells for induction of CTL may be important, because it is likely that several T-cell epitopes will be needed both for priming and to overcome allotype restriction. Nevertheless, inclusion of key epitopes may be desirable and the target region described by Takahashi *et al.*¹ is an important one because it is immunodominant for both neutralizing antibodies and CTL. Its drawback lies in its variability, although this might be overcome with appropriate cocktails¹⁵⁻¹⁷.

These results will stimulate a great deal of further research. But their ultimate value may depend on the extent to which

the alternative pathways can be used (see figure), particularly in vaccine development and immunotherapy. □

Dani P. Bolognesi is in the Department of Surgery, and the Department of Microbiology and Immunology, Box 2926, Duke University Medical Center, Durham, North Carolina 27710, USA.

1. Germain, R. *Nature* **322**, 687-689 (1986).
2. Takahashi, H. *et al.* *Nature* **344**, 873-875 (1990).
3. Staerz, U.D., Karasuyama, H. & Garner, A.M. *Nature* **329**, 449-451 (1987).
4. Deres, K. *et al.* *Nature* **342**, 561-564 (1989).
5. Nuchtern, J.G., Biddison, W.E. & Klausner, R.D. *Nature* **343**, 74-76 (1990).
6. Polydefkis, M. *et al.* *J. exp. Med.* **171**, 875-887 (1990).
7. Jin, Y. *et al.* *J. exp. Med.* **168**, 293-306 (1988).
8. Long, E.O. & Jacobson, S.J. *Immun.* **10**, 45-48 (1989).
9. Yewdell, J.W., Bennink, J.R. & Hosaka, Y. *Science* **239**, 637-639 (1988).
10. Sekaly, R.P. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 1209-1212 (1988).
11. Derosi, R.C. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **86**, 6353-6357 (1989).
12. Murphy-Corb, M. *et al.* *Science* **246**, 1293-1297 (1989).
13. Gibbs, C.J. *et al.* *Vth Int. Conf. on AIDS, Montreal, Canada (Abstr. C. Th. C.O. 46, 1989)*.
14. Stein, B.S. *et al.* *Cell* **49**, 659-668 (1987).
15. Putney, S.D. *et al.* *Vth International conference on AIDS, Montreal, Canada (Abstr. W.C.O. 18, 1989)*.
16. Zwart, G. *et al.* *Lancet* **ii**, 474 (1990).
17. Devash, Y. *et al.* *AIDS Res. hum. Retrovir.* (in the press).

In search of the single factor

Marvin R. Paule

EUKARYOTIC RNA polymerases are unable to recognize and transcribe from promoters at the beginning of genes without the aid of additional proteins, the general transcription factors. The three polymerases found in eukaryotic nuclei transcribe different sets of genes, and the number of factors associated with the initiation process increases as the variety of genes transcribed by the polymerase increases. Until now, the sole exception to the requirement for several transcription factors in eukaryotes was that of ribosomal RNA transcription by polymerase I from *Acanthamoeba*, for which a single ancillary protein (TIF-I) is required^{1,2}. But Kassavetis *et al.*³ now report that only one of the three factors involved in transcription mediated by polymerase III in yeast, TFIIB, is truly a transcription initiation factor. The others (TFIIA and TFIIC) are assembly factors responsible for loading the fundamental factor onto its site on the DNA. This observation means that the mechanisms described for rRNA transcription in *Acanthamoeba*, and now for transcription of yeast polymerase III genes, may be universal for eukaryotic transcription initiation — that is, that only one factor bound upstream of the transcription start site (dubbed +1) is needed to direct the polymerase to its binding site. The fundamental initiation factor can direct several rounds of initiation, and — most importantly — Kassavetis *et al.* show the addi-

tional factors which serve to load it on the promoter are dispensable once assembly of the stable transcription complex has been accomplished.

Acanthamoeba TIF-I assembles on the rRNA gene in the absence of additional factors. Yeast TFIIB, in contrast, cannot load independently onto the template: for 5S RNA genes, TFIIA and TFIIC must first bind to DNA, and, for transfer RNA genes, binding of TFIIC must precede the binding of TFIIB^{4,5}. Similarly, transcription of many vertebrate rRNA genes is stimulated by another factor, UBF, the cloning of which is reported by Jantzen and colleagues on page 830 of this issue. UBF increases binding of the TIF-I homologue; for example, in humans, at least tenfold stimulation of transcription occurs (see Fig. 3d of Jantzen *et al.*'s report). UBF from *Xenopus* cannot direct transcription alone, but it binds to the core promoter and to the rRNA gene enhancers (60/81 repeats), where it stimulates transcription, presumably by an effect on pre-initiation complex assembly⁷⁻⁹. These species differences between the transcription of rRNA in *Acanthamoeba* and vertebrates can be explained by the findings reported for yeast polymerase III.

By taking advantage of the extremely tight binding of yeast TFIIB once it has assembled on the template, Kassavetis *et al.*³ show that only TFIIB is required for transcription by yeast polymerase III.

They stripped TFIIA and TFIIC from the templates by using high salt concentration or heparin, and then isolated the DNA-protein complexes by size-exclusion chromatography. Only TFIIB complexes remained, and surprisingly, they retained the capacity for several rounds of transcription. The factors that had been stripped from the template were able to assemble fresh TFIIB on a new template; TFIIA and TFIIC are therefore assembly factors, and are not required for transcription initiation *per se*.

Pre-initiation complexes of tRNA and 5S RNA transcription in yeast^{3,5,10} and rRNA transcription in *Acanthamoeba*^{2,11,12} have been 'visualized' using footprinting techniques. Yeast TFIIB was found to protect the DNA template between about 10 and 40 base pairs upstream of +1 on 5S RNA and tRNA genes. The single *Acanthamoeba* rRNA transcription initiation factor (TIF-I) protected between about -12 and -70 base pairs, forming a stable complex which remained bound through several rounds of transcription (see figure). The two factors therefore form similar complexes upstream of the transcription start site.

The first footprints of a eukaryotic polymerase on a promoter demonstrated that *Acanthamoeba* RNA polymerase binds just downstream of TIF-I, protecting over 34 bp, to +18, from DNase I digestion². Replacement of the protected region with a variety of bacterial sequences showed that there are no DNA sequence-dependent contacts made by polymerase, but instead, the enzyme is positioned on the promoter by protein-protein contacts with TIF-I¹³. Similarly, polymerase III protects 23 bp (for 5S RNA genes) or 28 bp (for tRNA genes) of DNA just downstream of TFIIB. Earlier studies by Sakonju *et al.*¹⁴ suggest that the binding of polymerase to 5S RNA genes is also sequence-independent. To prove that the extended footprints are due to protection by polymerase and not to a conformational change in the previously bound factor, the polymerase I (refs 11, 13) and III (ref. 2) systems were supplied with a nucleotide mixture that allows the polymerase to make only a short RNA product. Addition of a mixture lacking GTP resulted in the polymerase stalling part way down the template. As predicted, the putative polymerase footprints moved part way down the DNA. Addition of all four nucleoside triphosphates resulted in total disappearance of the polymerase footprints.

Significantly, the *Acanthamoeba* TIF-I footprint remains unaltered during initiation, showing that TIF-I remains bound through several rounds of transcription^{11,12}. In the polymerase III systems, the TFIIB footprint was also unchanged after partial translocation of the polymerase down the template. Furthermore,

Shaky ground

ONE of the many hazards of earthquakes is the catastrophic 'liquefaction' of soil. The rapid succession of shocks shakes up the soil, mobilizes its water content, and sets it briefly flowing like a liquid. Buildings and structures previously supported on it are wrecked.

Daedalus is now taming this treacherous phenomenon. DREADCO's engineers are setting arrays of dynamite charges in soils ranging from sensitive sands and silts to well-consolidated clay. They then fire the charges in rapid succession, and examine the effects on test buildings and instruments scattered about the site. Their aim is to discover the frequency and amplitude of shock that best liquefies each type of soil.

Earthquake-levels of energy should not be needed. An optimized sequence of explosions, well tuned to the soil in question, should liquefy it very efficiently for a short while, with little other disturbance. Like a phased array, the charge pattern could be fired so as to launch its major disturbance in any desired direction, or even vertically downwards.

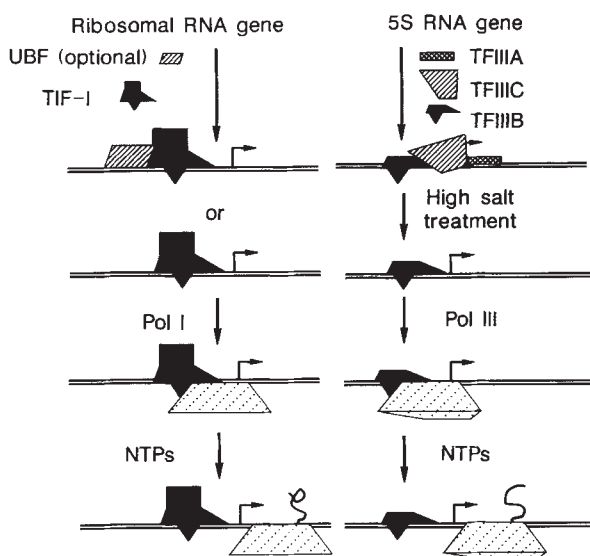
The way will be open for a whole new style of civil engineering, starting with elementary demolition. How elegant to liquefy the ground beneath some tower block or urban monstrosity, so that it simply sinks into the earth and vanishes without trace! Sadly, most monstrosities are not nearly heavy enough to sink completely into a cavity of liquefied soil. They would merely capsize and break up, leaving a flotsam of rubble.

Engineering construction could also be simplified. A complete array of piles could be sunk into liquefied soil in one operation, to be firmly held when it set again. Pipes and cables could be laid out around a site, and sunk to the required depth by a horizontally propagated wave of vibration, briefly liquefying the soil they lay on. Building foundations, erected on the surface, could be dropped to the damp-course by sudden liquefaction. Indeed, if designed for hydrostatic stability, an entire building might be constructed above ground, and 'launched' into the liquefied soil with speeches and champagne. It would then sink to its calculated Plimsoll line, submerging basement and foundations to their proper depth.

Daedalus feels that even agriculture could benefit. Dynamite has already been used for 'ploughing' in an experimental manner. But the complete fluidization of a field by a shallow omnidirectional shock-pattern should be far more efficient. Weeds would be overturned and ploughed in, pest larvae killed or severely shaken, deep nutrients mobilized, and the earth aerated, all in a few seconds. Even hidden archaeological remains might come floating to the surface.

David Jones

Initiation by eukaryotic RNA polymerases requires two classes of *trans*-acting protein factors, assembly (hatched) and initiation (black). Assembly factors mediate binding of the initiation factor to the promoter. For polymerase III (Pol III; right), assembly factors are obligatory, but can be removed after assembly of the initiation factor without effect on subsequent transcription (5S RNA is an example). For polymerase I (Pol I), the need for an assembly factor (UBF) is species-dependent. For both enzymes, the single transcription initiation factor (TIF-I for polymerase I, TFIIB for polymerase III) positions polymerase over the transcription start site (arrow) by protein-protein interaction. Polymerase loading is independent of DNA sequence. Initiation factors remain bound and can load additional polymerases. NTPs, nucleoside triphosphates.



Kassavetis *et al.* obtained nearly equal numbers of transcripts from complete and stripped templates³. Therefore, assembly factors are unnecessary for repetitive transcription initiation, even in systems that require them for pre-initiation complex formation, because the initiation factors do not cycle with each round of transcription.

What happens to assembly factors during transcription *in vivo*? Those bound upstream of the transcription start site, as on the rRNA genes, do not interfere with transcription. But assembly factors on 5S RNA and tRNA genes bind within the transcribed sequence, to the internal control region (ICR)¹⁵. Just how polymerase III is able to transcribe through the factors that bind to the ICR has been a puzzle. A clue comes from recent *in vivo* footprinting data¹⁶ suggesting that the assembly proteins do not remain tightly bound. This is in keeping with the latest discovery that the factors bound to the ICR can be removed without affecting subsequent initiations, showing that they are not transcription factors, but assembly factors. On the other hand, there is no evidence from studies *in vitro* showing that the assembly factors are released, so this remains an open question.

Are the same assembly factors used for all polymerase III genes? U6 small nuclear RNA, a component of spliceosomes, is transcribed by polymerase III, but it does not have the usual ICR sequences necessary for TFIIA or TFIIC binding¹⁷⁻¹⁹. Similar observations have been made for 7SL (ref. 20) and 7SK (ref. 21) genes. It seems that transcription of the U6 RNA gene does not require these assembly factors^{18,22}, although it still needs TFIIB. The U6 RNA gene, however, has promoter elements in the upstream

flanking region, and some of these function with transcription factors normally associated with RNA polymerase II^{17,23}. So alternative assembly factors may suffice for U6 RNA transcription.

It now remains to be seen whether the mechanism of transcription initiation that has now been established for polymerase I and III holds for polymerase II. □

Marvin R. Paule is in the Department of Biochemistry, Colorado State University, Fort Collins, Colorado 80523, USA.

1. Iida, C.T., Kownin, P. & Paule, M.R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1668-1672 (1985).
2. Bateman, E., Iida, C.T., Kownin, P. & Paule, M.R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8004-8008 (1985).
3. Kassavetis, G.A., Braun, B.R., Nguyen, L.H. & Gelduschek, E.P. *Cell* **60**, 235-245 (1990).
4. Bleker, J.J., Martin, P.L. & Roeder, R.G. *Cell* **40**, 119-127 (1985).
5. Braun, B.R., Riggs, D.L., Kassavetis, G.A. & Gelduschek, E.P. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2530-2534 (1989).
6. Jantzen, H.-M., Admon, A., Bell, S.P. & Tjian, R. *Nature* **344**, 830-836 (1990).
7. Pikaard, C.S., McStay, B., Schultz, M.C., Bell, S.P. & Reeder, R.H. *Genes Dev.* **3**, 1779-1788 (1988).
8. Bell, S.P., Pikaard, C.S., Reeder, R.H. & Tjian, R. *Cell* **59**, 489-497 (1989).
9. Pape, L.K., Windle, J.J., Mougey, E.B. & Sollner-Webb, B. *Molec. cell. Biol.* **9**, 5093-5104 (1990).
10. Kassavetis, G.A., Riggs, D.L., Negri, R., Nguyen, L.H. & Gelduschek, E.P. *Molec. cell. Biol.* **9**, 2551-2566 (1989).
11. Bateman, E. & Paule, M.R. *Cell* **47**, 445-450 (1986).
12. Bateman, E. & Paule, M.R. *Molec. cell. Biol.* **8**, 1940-1946 (1988).
13. Kownin, P., Bateman, E. & Paule, M.R. *Cell* **50**, 693-699 (1987).
14. Sakonju, S., Bogenhagen, D.F. & Brown, D.D. *Cell* **19**, 13-25 (1980).
15. Gelduschek, E.P. & Tocchini-Valentini, G.P. *A. Rev. Biochem.* **57**, 873-914 (1988).
16. Huibregtse, J.M. & Engelke, D.R. *Molec. cell. Biol.* **9**, 3244-3252 (1990).
17. Carbon, P.S. *et al. Cell* **51**, 71-79 (1987).
18. Reddy, R. *J. biol. Chem.* **263**, 15980-15984 (1988).
19. Kunkel, G.R., Maser, R.L., Calvet, J.P. & Pederson, T. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8575-8579 (1986).
20. Ulliu, E. & Weiner, A.M. *Nature* **318**, 371-374 (1985).
21. Murphy, S., di Liegro, C. & Melli, M. *Cell* **51**, 81-87 (1987).
22. Moenne, A. *et al. EMBO J.* **9**, 271-277 (1990).
23. Bark, C., Weller, J., Zabielski, J., Larson, L. & Pettersson, U. *Nature* **328**, 356-359 (1987).

Time for cigarette reclassification?

SIR—With so much ado about smoking one would think that official classifications of cigarettes would provide a reliable estimate of smoke intake from different brands. Alas, it is not so. Classifications to this day rely on analytical machines that puff cigarettes only at standard volume and frequency under standard laboratory conditions. In reality smokers puff and then inhale in amounts that vary from moment to moment, and draw into the lungs either much more or much less smoke than 'standard' smoking machines.

Today, surely, it is possible to develop an improved cigarette classification, because smoke intake can now be quantified by internalized markers. Of these, carbon monoxide, thiocyanate and others suffer from external interference whereas nicotine and cotinine are suitably specific. Virtually all available studies agree that plasma steady-state levels rise only around 2–4 per cent for nicotine and cotinine, compared with the increase of analytical machine yields measured by the smaller discriminant unit proffered in

would have aversive effects if personal thresholds are exceeded.

Depending on how they are smoked, nearly all cigarettes deliver enough nicotine to meet the fluctuating demands of any smoker, and therefore standard analytical yields of nicotine are meaningless as quantitative predictors of nicotine intake. Of course this is also true of tar yields. On the other hand, for any nicotine inhaled the corresponding amount of tar inhaled is defined by the tar-to-nicotine ratio of the smoke, a characteristic of each brand. A classification based on such ratios would then say what proportion of tar intake the smoker can expect from different brands.

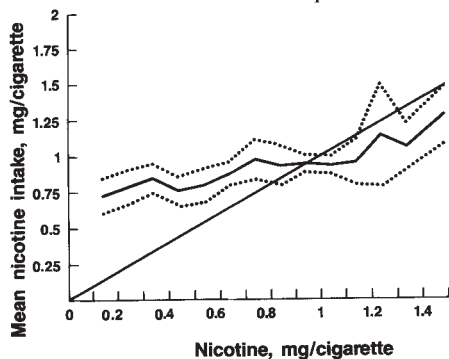
Tar-to-nicotine ratios change little within the range of smoking dynamics, the estimated variance being ± 10 –15 per cent^{8,9}. This would allow smoking machines to define and rank tar-to-nicotine ratios for different brands, especially after improvements in analytical sensitivity. Arguably, such ranking may have to account for the modest increase of intake when analytical yields increase. To this end, I suggest as a suitable formula: $\text{rank} = T(1+0.3N)/N$,

where T and N are mean standard machine yields of tar and nicotine in milligrams per cigarette, and 0.3 is a proportionality factor⁷. A continuous ranking scale of tar-to-nicotine ratios would encourage the development of cigarettes offering progressively lower tar and benign levels of nicotine, a trend already apparent in the market¹⁰ and officially considered as desirable¹¹.

GIO BATTÀ GORI

The Health Policy Center,
6704 Barr Road,
Bethesda,
Maryland 20816, USA

- Hill, P. & Marquardt, H. *Clin. pharmac. Ther.* **27**, 652–658 (1980).
- Russell, M.A.H. et al. *Br. Med. J.* **280**, 972–976 (1980).
- Benowitz, N.L. et al. *New Engl. J. Med.* **309**, 139–142 (1983).
- Ebert, R.V. et al. *J. Am. med. Ass.* **250**, 2940–2942 (1983).
- Gori, G.B. & Lynch, C.J. *Reg. Tox. Pharmac.* **5**, 314–326 (1985).
- Russell, M.A.H. et al. *J. epidem. Comm. Health* **40**, 80–85 (1986).
- Benowitz, N.L. et al. *J. Am. med. Ass.* **256**, 241–246 (1986).
- Browne, C.L., Keith, C.H. & Allen, R.E. *Beitrag zur Tabakforschung International* **10**, 81–90 (1980).
- Schlottzauer, W.S. & Chortyk, O.T. *J. analyt. Toxicol.* **7**, 92–95 (1983).
- Jarvis, M. & Russell, M.A.H. *Br. J. Addict.* **81**, 579–581 (1986).
- Wald, N. & Froggatt, P. (eds) *Nicotine, Smoking and the Low Tar Programme* (Oxford University Press, 1989).



Nicotine intake as a function of analytical cigarette yield. Yields determined by the FTC (Federal Trade Commission) standard method. Dotted lines, 95 per cent confidence intervals. Solid straight line, domain where intake would equal analytical yield (from ref. 5).

statutes and advertising, namely 0.1 milligrams for nicotine¹⁻⁷.

Based on pharmacokinetic considerations, we concluded⁵ that actual nicotine intake per cigarette increases by less than 3 per cent in relation to increasing analytical nicotine yields. Specifically, while such yields increase 1,400 per cent over the range 0.1–1.5 micrograms per cigarette, the corresponding mean nicotine intake per cigarette increases by only about 40 per cent (see figure).

It is thought that momentary behavioural and pharmacodynamic demand for nicotine modulates puffing and inhalation. At the same time, the observed daily steady-state plasma nicotine levels show that cumulative demand may be limited at quasi-stable ceilings characteristic of each individual. It is conceivable that nicotine

Pseudo-creation of the Big Bang

SIR—John Maddox¹ has rejected the Big Bang model of the Universe as “philosophically unacceptable” and Jean-Marc Lévy-Leblond² has replied, “it need not be as ‘philosophically unacceptable’ as he contends.” But Maddox expects, and Lévy-Leblond allows, that scientific evidence will turn out to justify the abandonment of the model.

I claim that, insofar as both its classical and quantum versions become unacceptable, they will do so only on scientific, rather than on philosophical grounds. For example, the very recently discovered “great wall” and “great attractor”, the so-called “dark matter”, the newly observed most distant and oldest quasars³ and the role played by plasma in cosmic evolution⁴ pose a theoretical challenge that the Big Bang framework may, in due course, be unable to meet.

But there are basic logical flaws in the assumptions that prompted Maddox’s philosophical rejection of the cataclysmic cosmic scenario as well as in Lévy-Leblond’s claim that his linear time has enhanced the conceptual palatability of the Big Bang. They either beg the question or illicitly impose *a priori* demands on physical cosmology.

The “philosophical difficulty” that Maddox sees in the Big Bang model is “that an important issue, that of the ultimate origin of our world, cannot be discussed”. His argument is predicated on

the occurrence of the cosmic cataclysm at a “well-defined instant” t_0 which had no temporal predecessor. But as Torretti has explained⁵, general relativity theory does not countenance t_0 as a bona fide instant of the Big Bang space-time.

Yet the alleged philosophical difficulty would be spurious, I contend, even if the model did feature a time-interval that is closed, rather than open, for $t=t_0$. By hypothesis, there simply did not exist any instants before it. But precisely this total absence of times earlier than t_0 also rules out the very existence of an earlier cause of any event that does occur at the hypothesized instant t_0 . Hence, if the Big Bang is taken to have occurred at the putative t_0 , that initial event is causally *sui generis*. It just cannot have any cause at all in the universe of the given model, nor, of course, can that paramount occurrence be the effect of any prior cause.

Therefore, it just begs the question to insist, as Maddox does peremptorily, on characterizing the supposedly initial event as “an effect”. By thus illicitly requiring the existence of an earlier cause within the assumed model, this characterization also requires the existence of at least one instant before t_0 after all, which saddles the cosmological model with a temporal inconsistency. Maddox objects that, *qua purported effect*, the Big Bang “is an effect whose cause cannot be identified or even discussed”, but the elusiveness of the

phantom earlier cause is due to its sheer non-existence in the face of the demand that it must exist nonetheless.

Evidently, it is imperative to distinguish the sound question "Does the universe have a temporal beginning?" from the quite different alleged problem, "If the universe did have a bounded past of finite duration, what was the cause of its initial event at t_0 ?" Maddox takes the latter query to be the "important issue . . . of the ultimate origin of our world." But, absent the phantom cause or "ultimate origin", the craving for it — and for the causal explanation of the classical Big Bang at the putative instant t_0 — is a seductive pseudo-problem, rather than a riddle eluding scientific solution. A question cannot be regarded as a well-posed challenge to a theory merely because the questioner finds it psychologically insistent, and finds the answer to it elusive after introducing a temporal inconsistency into the theory *a priori*.

Though the Big Bang model is beset by some scientific liabilities, it is not vitiated by the purported difficulty. *A fortiori*, there is no warrant at all for Maddox's contention that "creationists . . . have ample justification in the doctrine of the Big Bang." Indeed, the putative cosmological model poses no challenge whatever to atheism⁶. It is the *a priori* philosophical aversion for a bounded, metrically finite past that plays into the hands of the creationists. This aversion goes back to Aristotle, who indirectly ruled out the supposed Big Bang model *a priori* by deeming an instant without temporal predecessor to be inconceivable (*Physics*, Book VIII, 251 b). But such a moment is quite conceivable, and the verdict as to its actual existence must be reserved to the mathematics of our best physical theories.

As noted by Lévy-Leblond², general relativity excludes t_0 from the set of bona fide physical instants. Assuming also that there is no final future crunch, he points out that the cosmic time-interval of the Robertson-Walker spacetime metric is open in both directions ($t_0 < t < \infty$) but patently does not extend before t_0 . Relying on an *a priori* prejudice, however, Lévy-Leblond finds it philosophically discomfiting that, even though the past on this open interval is ordinarily unbounded, the age of the Big Bang universe in the given metric is finite, presumably somewhere between 12 and 20 billion years.

Citing Misner⁷, he aims "to send back the birth of the Universe [metrically] to (minus) infinity where, or rather when, it seems to belong." Hence he was pleased to introduce his new linear time metric, which confers an infinite duration on the ordinarily unbounded past of the Big Bang universe.

True, this alternative time-metrization

is quite legitimate. Interestingly, it is physically realized, as Lévy-Leblond explains, by a clock geared to the expansion of the universe. But just why must the birth of the universe "belong" to a past of metrically infinite duration? Since the philosophical malaise experienced by those who shrink *a priori* from a metrically finite age of the Universe is baseless, the ability to allay this discomfiture does not add merit to the remetrization. *Pace* Maddox and Lévy-Leblond, the Big Bang model featuring a finite age of the Universe on the standard cosmic timescale is not philosophically disquieting at all.

In one of the versions of current quantum cosmology⁸, the Big Bang is no longer held to comprise all of the earliest instants; instead, the inflationary expansion is preceded by two so-called "vacuum" states in a metrically finite past. Within that theoretical framework, it is no longer misguided to ask "what caused the Big Bang?" Given the model's random quantum fluctuations in its "false" vacuum, the general theory of relativity tells us why there was an inflationary expansion. Here as well, the model lends no support to the doctrine of divine creation *ex nihilo*^{6,9}, but that teaching lingers on in the misleading overtones of the terms "creation" and "nothing". The former connotes the operation of a creator or external causal agency, and the latter a completely unstructured state.

Alas, the recent literature on some versions of quantum cosmology contain inappropriate uses of these locutions which may suggest that this theory abets

creationism. For example, such physicists as Hartle and Hawking¹⁰ and Vilenkin¹¹ speak misleadingly of certain primordial physical states as "nothing", even though these states are avowedly only "a realm of unrestrained quantum gravity," which is "a state with no classical space-time"¹¹.

Recently, plasma cosmology⁴ has posed a major challenge to the gravity-dominated Big Bang models by assigning a critical cosmic role to hot electrically charged gases. The plasma model evolves without any beginning, being metrically infinite in both time-directions on the standard cosmic scale. But this feature does not make the plasma universe philosophically preferable to any of its Big Bang rivals. The merits of their competing claims turn instead on their scientific credentials, which include the adequacy with which they fit observational findings.

ADOLF GRÜNBAUM

*Humanities and Social Sciences,
California Institute of Technology,
Pasadena,
California 91125, USA*

1. Maddox, J. *Nature* **340**, 425 (1989).
2. Lévy-Leblond, J.-M. *Nature* **342**, 23 (1989); *Am. J. Phys.* **58**, 156–159 (1990).
3. Wilford, N.W. *The New York Times*, Section E, p. 5. 21 January (1990).
4. Peratt, A.L. *The Sciences* January/February 24–32 (1990).
5. Torretti, R. in *Relativity and Geometry*, 210–219 (Pergamon, New York; 1983).
6. Grünbaum, A. *Phil. Sci.* **56**, 373 (1989).
7. Misner, C.W. *Phys. Rev.* **136**, 1328 (1969).
8. Weisskopf, V. *Bull. Am. Acad. Arts Sci.* **42**, 22 (1989).
9. Grünbaum, A. *Proc. Int. Congr. "Truth in Science," Accad. Naz. dei Lincei*, (Rome, Italy 1989).
10. Hartle, J.B. & Hawking, S.W. *Phys. Rev.* **D28**, 2960 (1983).
11. Vilenkin, A. *Phys. Rev.* **D27**, 2848 (1983).

A puzzling pulsar companion

SIR—In his recent News and Views article¹, Lyne expressed concern with the *a priori* improbability of encountering a binary pulsar with the characteristics derived by Ables *et al.*² from their observations of PSR0021–72A. Lyne tried to solve the problem by seeking other interpretations of the observations. But, in fact, these are all excluded by the observations.

First, a note of caution is appropriate: if a population ("the PSR0021–72A-like sources") has only one known member, its mean properties have infinite errors, as do extrapolations based on them. In any case, Ables *et al.* do consider the question of whether the surprisingly low orbital inclination implies that there are many more binary millisecond pulsars in 47 Tuc, and state that this is probably not the case, because any other pulsar with a similar period and flux level would have shown up in their power spectra, whatever its inclination.

The possibility that the companion is a small planet was considered by one of us (R.A.M.J.W.)³ — and probably by many

others — with the conclusion that only degenerate companions are allowed. To be more explicit, Jupiter would be far too large to fit into its Roche lobe, but planets composed of very heavy elements (for equation of state see ref. 4) would both fit into their Roche lobes and yield the observed value of the apsidal motion. Examples are a 0.016 $M_\odot$ iron planet ($R = 0.028 R_\odot$, orbital inclination $i = 14^\circ$) or a 0.009 $M_\odot$ lead planet. But those more, not less, exotic alternatives are ruled out by the observed amplitude of the Einstein delay, 50–80 times that provided by planets of these masses.

Lyne also suggests that the observed behaviour of pulse arrival times may be attributed to an intrinsic phase modulation arising from a precession or a torsional oscillation of the neutron star. To match the amplitude of the delay in pulse arrival times, such a phase modulation would be required to have an amplitude greater than π . However, a precession of the neutron star can never produce a phase modulation with amplitude larger than $\pi/2$ (ref. 5). A

torsional oscillation of the star can, in principle, produce the required amplitude of phase modulation, but it would be nearly impossible to reproduce the apsidal motion and the Einstein delay. Moreover, this object would still be very special, since no other radio pulsar is known to exhibit a similar behaviour.

In short, it requires quite a cosmic conspiracy for any physical mechanism (other than the motion in a binary system) to produce a pulsar signal with a phase modulation that accurately mimics an eccentric binary orbit with two relativistic corrections.

D. BHATTACHARYA
R. A. M. J. WIJERS

Astronomical Institute,
University of Amsterdam,
Roetersstraat 15, NL-1018 WB,
The Netherlands

1. Lyne A. *Nature* **342**, 128–129 (1989).
2. Ables, J. G. et al. *Nature* **342**, 158–161 (1989).
3. Wijers, R. A. M. J. *Astr. Astrophys.* **209**, L1–L4 (1989).
4. Zeplosky, H. S. & Salpeter, E. E. *Astrophys. J.* **158**, 809–813 (1969).
5. Nelson, R. W., Finn, L. S. & Wasserman, I. *Astrophys. J.* **348**, 226–231 (1990).

Early tetrapod hearing

SIR—I should like to respond to the comments in Scientific Correspondence^{1,2} on my recent Letter on the stapes of the early tetrapod *Acanthostega*³. This bone is pierced by a stapedia foramen, clearly marking it as a stapes rather than a fish hyomandibula⁴: a more certain basis for distinction than inferred function¹.

Second, neither the hyomandibula of osteolepiforms (the group of lobe-finned fishes presumed ancestral to tetrapods) nor the stapes of *Acanthostega* have direct connections with the jaws. The suggestion² that this decoupling in *Acanthostega* is a major difference from the osteolepiform condition is therefore untrue.

Third, we need not accept that osteolepiforms had a tympanic membrane², with the implication that an open spiracle is incompatible with an air-filled spiracular pouch. Spiracles can be closed, and a pouch with a functional spiracle might also act as a middle ear cavity. The existence of a basilar papilla in the extant coelacanth *Latimeria*² suggests that an auditory function for the inner ear is a primitive feature of lobe-finned fishes, in association with an air-filled spiracular pouch with or without a membranous closure.

JENNIFER A. CLACK

University Museum of Zoology,
University of Cambridge,
Downing Street,
Cambridge CB2 3EJK, UK

1. Gottfried, M. D. & Foreman, B. *Nature* **344**, 116 (1990).
2. Fritsch, B. *Nature* **344**, 116 (1990).
3. Clack, J. A. *Nature* **342**, 425–427 (1989).
4. Smithson, T. R. & Thomson, K. S. *Zool. J. Linn. Soc.* **74**, 93–103 (1982).

When is random random?

SIR—The intelligentsia is not on speaking terms with itself. Ochs¹ correctly cautions against sloppy use of the term 'random', but seems unaware of the work of mathematicians on this problem over many years. If each event in a set is selected with equal probability, say the toss of a fair coin, the selection is indeed random. But the probability of a 7 in a fair throw of two dice is not the same as that of a 2, yet the process is clearly to be considered random. Thus there is more randomness in a sequence of heads or tails from the toss of a fair coin than in a sequence of throws of two dice. Clearly a more sophisticated definition of randomness is required than one finds in a dictionary.

Suppose we wish to send to our little green friends on a planet near Arcturus the value of π . We could start sending the digital sequence in the decimal representation of π , but that would take an infinite time and an infinite amount of information, for there is no pattern among these digits. But π is a transcendental number, and we could send one of the algorithms from which it is calculated, not necessarily the shortest, letting our little green friends do the rest. Thus it is clear that π and all other transcendental numbers have a finite amount of information, namely, that in the algorithm by which they are calculated.

In the case of a sequence of heads and tails generated by the tossing of fair coin, there is no algorithm which can shorten the sequence. We must simply send the sequence itself. Such random sequences are known as Bernoulli chains or strings. On the other hand, the probability distribution in the tossing of two fair dice may be used to shorten the sequence of numbers generated. This is because some information is lost when the numbers on the faces of the two dice are added and

only the sum is recorded. (We cannot know from the sum, especially in the case of 7, the exact faces which were uppermost.)

Thus we can measure the amount of randomness in a sequence by the length of the algorithm, in bits by which we can describe it to our little green friends. This fact emerged from the initially independent work of Solomonoff², Kolmogorov³, Chaitin⁴ and Martin-Löf⁵ on the foundations of probability theory. Sequences which cannot be described by an algorithm shorter than the sequence itself have the greatest measure of randomness. Because any computable number can be described by an algorithm which is finite in length, no computable number is random⁶. It is impossible to prove that any given sequence was generated by a random process.

The current state of the theory is due to Chaitin⁶. I suggest that this literature be consulted by those who are interested in the proper use of the term 'random'. Ochs¹ repeats a common mistake in that he attempts to squeeze meaning out of words such as 'chaotic, unpredictable, uncertain, arbitrary, undetermined'. 'Arbitrary' and 'undetermined' do not mean 'random'. Words are merely names of mathematical functions and take their meaning from the mathematics, not the other way around.

HUBERT P. YOCKEY

1507 Balmoral Drive,
Bel Air,
Maryland 21014, USA

1. Ochs, T. L. *Nature* **343**, 303 (1990).
2. Solomonoff, R. J. *Inform. Control* **7**, 1–22, 224–254 (1964).
3. Kolmogorov, A. N. in *Problems of Information Transmission* (Vol. 1) (IEEE 1965).
4. Chaitin, G. J. *J. ACM* **13**, 547–569 (1966).
5. Martin-Löf, P. *Inform. Control* **9**, 602–619 (1966).
6. Chaitin, G. J. in *Information, Randomness and Incompleteness*, (World Scientific, Singapore, 1987).

Cold shock and DNA binding

SIR—Goldstein *et al.*¹ have recently described a gene from *Escherichia coli* that is induced in response to low temperature, giving rise to a cold-shock protein (csp) CS7.4. No similarity between CS7.4 and any other protein was noted by these authors. But when this sequence is compared with GenBank v62 using the program SEQFT², a striking similarity with a conserved region of human DNA-binding proteins is revealed. The human proteins DbpA,

DbpB (ref. 3) and YB-1 (ref. 4) share a region of conserved sequence which closely matches the entire deduced sequence of CS7.4.

The function of CS7.4 is unknown, with suggestions varying from anti-freeze protein to a role in initiating translation. It has also been suggested that the protein is autoregulatory, interacting with its own gene promoter¹. DbpA and DbpB are DNA-binding proteins of unknown specificity³ but YB-1 has

```

csp  MSGKMTGIVKWFNADKGFETIPDDGSKDVFVHFAIQNDG---Y-KSLDEGQKVSFTIESGAKGPAAGNVTSL  1-70
dbpa  LATKVLGTVKWFNVRNGYGFINRNDTKEDVFVHQTAKKNNPRKYLRSVGDGETVFEDVVEGEGKGAEEANVTGP  157-230
dbpb  IATKVLGTVKWFNVRNGYGFINRNDTKEDVFVHQTAKKNNPRKYLRSVGDGETVFEDVVEGEGKGAEEANVTGP  95-168
yb1   IATKVLGTVKWFNVRNGYGFINRNDTKEDVFVHQTAKKNNPRKYLRSVGDGETVFEDVVEGEGKGAEEANVTGP  55-127

```

Alignment of deduced amino-acid sequences of CS7.4 (csp)¹, DbpA and B³ and YB-1⁴. Sequence positions are indicated. Residues which are identical between the prokaryotic and eukaryotic proteins are marked with an asterisk; conservative changes are marked with a dot.

been shown to bind to the CCAAT-containing Y box of HLA class II genes⁴. Conceivably, the conserved sequence in all these proteins represents a novel DNA recognition motif, one which is found in both prokaryotes and eukaryotes.

1. Goldstein, J., Pollitt, N.S. & Inouye, M. *Proc. natn. Acad. Sci. U.S.A.* **87**, 283–287 (1990).
2. Kanehisa, M. *IDEAS Users Manual* (FCRF, Frederick, Maryland, 1986).
3. Sakura, H. *et al. Gene* **73**, 499–507 (1988).
4. Didier, D.K. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 7322–7326 (1988).
5. Garnier, J., Osguthorpe, D.J. & Robson, B. *J. molec. Biol.* **120**, 97–120 (1978).

Secondary structure predictions⁵ for these sequences are somewhat inconclusive but it is interesting that there are six invariant glycine residues, suggesting that critical turns or bends are important features of the tertiary structure.

GRAEME WISTOW

*Laboratory of Molecular and
National Eye Institute,
Developmental Biology,
National Institutes of Health,
Bethesda,
Maryland 20892, USA*

Radiation and exposure rate

SIR—Substantial evidence has recently emerged that the coefficients relating the risk of lung cancer to exposure to radiation from radon in mines differ with the rate of exposure, and that the relative risk increases with decreasing exposure rate. This may be due to a genuine exposure rate effect or, possibly, to an effect of total dose.

The first evidence came from studies^{1,2} of the workforces at two uranium mines in Canada, operated by Eldorado Resources Ltd^{1,2}. Surveys at each mine allowed the lung cancer risk to be related to individuals' exposure, in working level months (WLM). At Beverlodge (Saskatchewan) and Port Radium (Northwest Territories), the increase in relative risk with exposure was found to be 3.28 per cent per WLM (95 per cent confidence interval 2.08, 4.48) and 0.27 per cent (95 per cent confidence interval 0.11, 0.43), respectively.

The difference between the absolute risks is equally striking, being 20.8 per 10⁶ person-years per WLM at Beverlodge (95 per cent confidence interval 11.8, 29.8) and 3.10 at Port Radium (95 per cent confidence interval 1.89, 4.32). The chief difference between the two mines seems to be the average exposure rate, estimated at 5 WLM a year at Beverlodge and 109 at Port Radium. Further evidence that risk is related to exposure rate has come from three sources.

First, Hornung and Meinhardt³ have shown an inverse exposure-rate effect among uranium miners on the Colorado Plateau. They conclude that, among miners with equivalent cumulative exposure, the risk of lung cancer is greater at lower levels of exposure for longer periods of time. Moreover, this effect varies with cumulative exposure. For miners of similar age and smoking status, and with similar cumulative exposure, Hornung and Meinhardt find that a 90 per cent reduction in exposure rate is associated with a 58 per cent increase in risk for miners with cumulative exposure in the range 0–834 WLM, but that a similar reduction of exposure rate is associated with only a 10 per cent increase in risk if the cumulative exposure is in the range 834–10,000 WLM.

Second, the BEIR IV committee⁴ has estimated risks for four cohorts of miners (see table, which includes data from Port Radium not available to the committee). The estimates shown are specific for ages 55–64 and a period 5–14 years after exposure, thus avoiding the complications of differences of age structure between the cohorts and different follow-up intervals. Again, it seems that lower exposure rates are associated with the higher risks. Third, a similar effect among Czech miners has been briefly reported⁵.

By themselves, these studies do not exclude the possibility that the observed

association reflects confounding between exposure rate and total exposure: workers with lower exposure rates tend to have lower cumulative exposure. But if this were the case, one would expect that the slope of the dose–response relationship at high exposure rates (Colorado, Port Radium) would be greater at lower than at higher total dose. Such an effect is indeed apparent in the Colorado data, for which reason the BEIR IV committee used only cumulative lagged exposures of less than 2,000 WLM in its analysis. Below this level, the data from Colorado and Port Radium are inconclusive.

Whatever the explanation, the practical consequences for evaluation of the risks of exposure to indoor radon are the same. Because the exposure rate is low (about 0.1 WLM per year in Britain, for example), estimates of risk are more appropriately based on studies such as those at Beverlodge and Malmberget rather than on all the available data. Relative risks thus estimated will be greater than the BEIR IV estimate of 2.5 per cent per WLM by between 50 and 100 per cent (the BEIR IV excess relative risk at Beverlodge and Malmberget over the overall BEIR IV estimate; see table).

If there should be a further increase of relative risk with decreasing dose rate below about 5 WLM a year, even these increased estimates may be too low. The only study so far to have reported⁶ direct estimates of the effect at indoor levels of exposure to radon suggests an increase in relative risk of about 3.4 per cent per WLM, but this estimate is imprecise (95 per cent confidence interval 0.0, 8.0). Estimation of the effect by extrapolation is complicated by the interaction of radon exposure and smoking, which may not be multiplicative⁷, but for non-smokers an upper limit can be derived from the observation that their lifetime risk of lung cancer of about 0.6 per cent is, with few exceptions, remarkably constant in all populations for whom data are available, despite a variation of at least twofold in the average national exposure to radon indoors.

SARAH C. DARBY

RICHARD DOLL

*Imperial Cancer Research Fund,
Cancer Epidemiology Unit,
University of Oxford,
Radcliffe Infirmary,
Oxford OX2 6HE, UK*

RELATIVE RISKS AND AVERAGE EXPOSURE RATES IN FIVE COHORTS OF MINERS EXPOSED TO RADON

Cohort	Average exposure rate (WLM per year)	Relative risk (per cent per WLM)	
		Original*	BEIR IV†
Beverlodge	5	3.3	5.1
Malmberget	5	3.6	3.6
Ontario	10 (approx.)	1.3	1.8
Colorado	124	‡	0.9§
Port Radium	109	0.27	—

* Estimate given by original authors.

† BEIR IV estimate, based on internal comparison, for age 55–64, 5–14 years after exposure.

‡ Original paper gives estimates only after allowing for exposure rate effect and are thus not directly comparable.

§ Based on cumulative lagged exposures of less than 2,000 WLM only.

1. Howe, G.R., Nair, R.C., Newcombe, H.B., Miller, A.B. & Abbott, J.D. *J. natn. Cancer Inst.* **77**, 357–362 (1986).
2. Howe, G.R. *et al. J. natn. Cancer Inst.* **79**, 1255–1230 (1987).
3. Hornung, R.W. & Meinhardt, T.J. *Health Phys.* **52**, 417–430 (1987).
4. *Committee on the Biological Effects of Ionizing Radiations (BEIR IV)* (National Academy Press, Washington, DC, 1988).
5. Sevc, J., Kunz, E. & Tomasek, L. *Health Phys.* **54**, 27–46 (1988).
6. *New Jersey State Department of Health Technical Report-Phase I* (1989).
7. Lubin, J.H. *Yale J. biol. Med.* **61**, 195–214 (1988).

The Oort cloud

Paul R. Weissman

Although the outermost planet, Pluto, is 6×10^9 km from the Sun, the Sun's gravitational sphere of influence extends much further, out to $\sim 2 \times 10^{13}$ km. This space is occupied by the Oort cloud, comprising 10^{12} – 10^{13} cometary nuclei, formed in the primordial solar nebula. Observations and computer modelling have contributed to a detailed understanding of the structure and dynamics of the cloud, which is thought to be the source of the long-period comets and possibly comet showers.

THE huge spherical cloud of comets surrounding the planetary system was first suggested in 1950 by Dutch astronomer Jan Oort¹ to explain the unusual distribution of orbital energies for the observed long-period (LP) comets (Fig. 1a) as a function of the original inverse semi-major axis, $1/a_0$, which is equivalent to orbital energy². The original orbit is that of the comet integrated backwards in time to before it entered the planetary system, and referred to the barycentre of the Solar System. The $1/a_0$ distribution consists of: (1) a sharp spike of comets at near-zero but bound energies, representing orbits with semi-major axes between 10^4 AU and infinity (1 astronomical unit (AU) is equal to the Earth–Sun distance, ~ 150 million km); (2) a low continuous distribution of more tightly bound, less eccentric orbits; and (3) a few apparently hyperbolic orbits. Earlier dynamical calculations³ had shown that Jupiter would perturb the orbits of comets passing through the planetary system so as to randomly spread them in energy, giving rise to the low continuous distribution. But what then was the explanation for the spike of comets?

Oort recognized that the spike had to be the source of the long-period comets, a vast spherical cloud of comets at distances greater than 10^4 AU from the Sun, but still gravitationally bound to it. Oort showed that comets in the cloud are so far from the Sun that perturbations from random passing stars can change their orbits and occasionally send the comets back into the planetary system. On their first pass through the planetary system, Jupiter's random perturbations eject roughly half the 'new' comets to interstellar space, and capture the other half to more tightly bound, less eccentric orbits. Only 5% of the new comets are returned to Oort cloud distances⁴. On subsequent returns the comets continue to 'random-walk' in orbital energy until they are either ejected, collide with a planet or the Sun, or are destroyed by one of several poorly understood physical mechanisms. The few 'hyperbolic' comets in Fig. 1a are probably the result of small errors in orbit determination or of unmodelled non-gravitational forces resulting from jetting of volatiles on the surfaces of the cometary nuclei⁵, which make the orbits seem more eccentric than they actually are.

Oort's accomplishment in defining the source of the long-period comets is even more impressive when one considers that it was based on only 19 well determined cometary orbits, compared with the ~ 190 quality orbits in Fig. 1a. Further analysis of the observed orbits² shows that the average long-period comet entering the planetary system comes from an aphelion distance of 4.3×10^4 AU.

Some have pointed out that Öpik⁶ may have anticipated Oort's work in his studies, 18 years earlier, of the effects of stellar perturbations on distant meteor and comet orbits. Öpik did suggest that stellar perturbations would increase the perihelia of comets, resulting in a cloud of objects surrounding the Solar System. But he specifically rejected the idea that comets coming from the cloud could ever be observed because he did not recognize that stellar perturbations would also cause some orbits to diffuse back into the planetary region. Öpik concluded that the observed long-period comets came from aphelion distances

of only about 1.5 – 2.0×10^3 AU. Although Öpik's 1932 paper was a pioneering work on stellar perturbations, it did not identify the cometary cloud as the source of the long-period comets or relate the observed orbits to the dynamical theory.

Interestingly, speculations by Halley⁷ in his classic work on comets could be interpreted as implying a distant comet cloud.

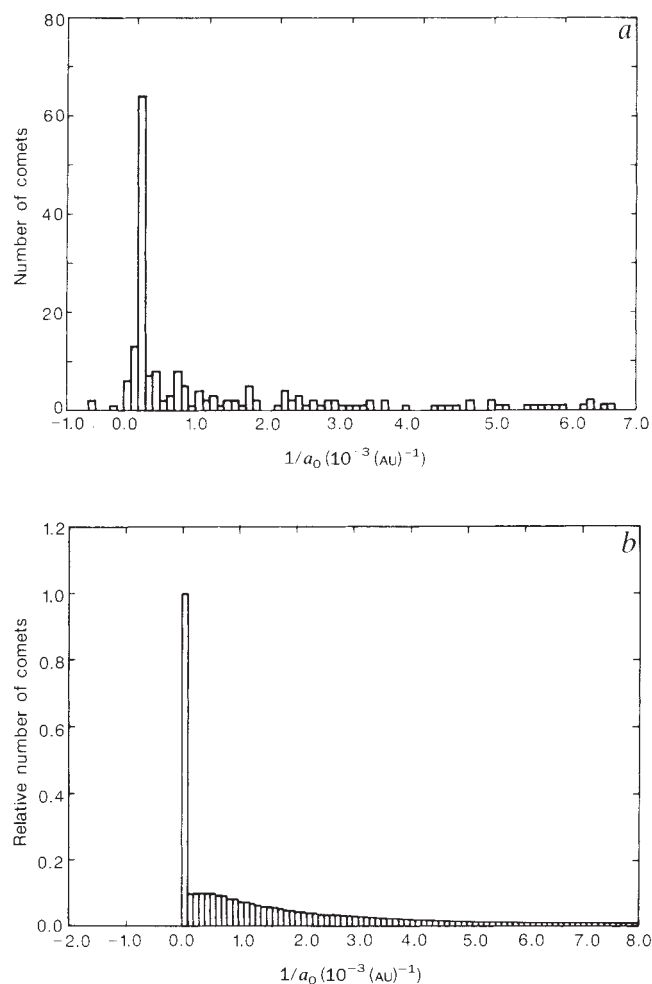


FIG. 1 a, Distribution of original inverse semi-major axes for the observed long-period comets. The sharp spike of comets at near-zero but bound energies are the dynamically 'new' comets from the Oort cloud. The few hyperbolic orbits are probably the result of errors in their determinations. Orbital data from Marsden *et al.*². b, Estimated distribution of inverse semi-major axes of 10^5 comets based on a Monte Carlo simulation model⁴ that includes the effects of planetary, stellar and non-gravitational perturbations, and physical loss due to collisions, random disruption and loss of all volatiles.

Halley was able to fit only parabolic elements to the 24 sets of comet observations that he studied, but he argued that the orbits would prove to be elliptic, writing, "For so their Number will be determinate and, perhaps, not so very great. Besides, the Space between the Sun and the fix'd Stars is so immense that there is Room enough for a Comet to revolve, tho' the Period of its Revolution be vastly long".

The dynamical evolution of comets in the Oort cloud can be simulated using a computer-based Monte Carlo model. The results of one such simulation⁴ for 10^5 hypothetical comets are shown in Fig. 1b, where the comets are perturbed by the major planets, random passing stars and non-gravitational forces, and are removed by collisions, random disruption (splitting) and loss of all volatiles. A fairly good match to the observed distribution in Fig. 1a is obtained. By 'tuning' such a model to improve the fit, one can gain some insight into the possible physical and dynamical loss mechanisms.

These simulations show that 65% of all long-period comets passing through the region of the inner planets are dynamically ejected from the Solar System on hyperbolic orbits, 27% are randomly disrupted (10% on the first perihelion passage) and the remainder are lost by a variety of processes such as loss of all volatiles and collision with the Sun and planets. The average ejection velocity is 0.6 km s^{-1} . The average long-period comet with perihelion $< 4 \text{ AU}$ makes five passages through the planetary region before arriving at some end-state, with a mean lifetime of $6 \times 10^5 \text{ yr}$ between the first and last passages.

To account for the observed flux of dynamically 'new' long-period comets, Oort estimated that the population of the cometary cloud was 1.9×10^{11} objects. More recent dynamical models^{8,9} have produced somewhat higher estimates, $\sim 1\text{--}2 \times 10^{12}$ comets. These result in part from higher estimates of the flux of long-period comets (brighter than absolute magnitude $H_{10} = 11$) through the planetary system, after correction for observational selection effects¹⁰, and in part from a recognition of the role of the jovian planets in blocking the diffusion of cometary orbits into the planetary region¹¹. Comets perturbed inward to perihelia near the orbits of Jupiter and Saturn will probably be hyperbolically ejected before they can diffuse to smaller perihelia and be observed. Thus, the region of the terrestrial planets is undersupplied in long-period comets.

Oort cloud perturbers

Since first proposed in 1950, Oort's vision of a cometary cloud gently stirred by perturbations from distant passing stars has evolved considerably. Other sources of perturbation have been recognized: giant molecular clouds (GMCs) in the galaxy^{12,13} (unknown before 1970), and the galactic gravitational field itself, in particular the tidal field of the galactic disk¹⁴⁻¹⁶. GMC encounters are rare, occurring with a mean interval of $3 \times 10^8 \text{ yr}$, but result in large perturbations on the orbits in the Oort cloud. The galactic field also sets limits on the outer dimensions of the Oort cloud. The cloud is a prolate spheroid with the long axis oriented toward the galactic nucleus^{17,18}. Maximum semi-major axes are $\sim 10^5 \text{ AU}$ for direct orbits (relative to galactic rotation) oriented along the radius vector to the galactic nucleus, decreasing to $\sim 8 \times 10^4 \text{ AU}$ for orbits perpendicular to the galactic radius vector, and increasing to $1.2 \times 10^5 \text{ AU}$ for retrograde orbits (those opposite to the direction of galactic rotation).

In addition, random stars occasionally pass directly through the Oort cloud, ejecting a substantial number of the comets and severely perturbing the orbits of others¹⁹. The passage of a star drills a narrow tunnel through the Oort cloud by ejecting all comets within a radius of $\sim 470 \text{ AU}$, for a $1\text{-}M_{\odot}$ star passing at a distance of $5 \times 10^4 \text{ AU}$ from the Sun and at a velocity of 20 km s^{-1} (refs 20, 21). Over the history of the Solar System $\sim 5.4 \times 10^3$ stars have passed within 10^5 AU of the Sun, ejecting $\sim 10\%$ of the Oort cloud population.

It is now recognized that the galactic disk is the major perturber on the Oort cloud in terms of bringing new comets into the

planetary system, although stars and GMCs still have an important role in repeatedly randomizing the cometary orbits. Galactic tidal perturbations peak for orbits with their lines of apsides (the line connecting perihelion and aphelion) at galactic latitudes of $\pm 45^\circ$ and go to zero at the galactic equator and poles^{14,16}. The distribution of latitudes of aphelion directions of the observed long-period comets mimics that dependence, as shown in Fig. 2 (ref. 22). Although the lack of discovered comets near the galactic equator could be the result of observational selection (confusion with galactic nebulae), the lack of comets near the poles seems to confirm the importance of the influence of the galactic disk on the Oort cloud.

Because the galactic tide changes the cometary perihelia in a regular fashion¹⁶ rather than randomly as do stellar perturbations, it is easier to overcome the dynamical barrier that Jupiter and Saturn present to cometary diffusion into the region of the inner planets. Thus the comets are brought into the observable region more efficiently. As a result, estimates of the population of the Oort cloud should decrease, although detailed estimates have yet to be made. One Monte Carlo simulation indicates that the decrease might be as much as a factor of four²³, implying a population of only $\sim 3\text{--}5 \times 10^{11}$ comets.

The inner cloud

As a result of the better understanding of the competing roles of Oort cloud perturbers, it is now estimated that the mean dynamical lifetime of comets in the Oort cloud is only $\sim 60\%$ the age of the Solar System²⁴ (some estimates are even shorter²⁵), which means that the cloud must somehow be replenished. Replenishment can result by capture of comets either from interstellar space²⁶ or from a more populous inner Oort cloud reservoir, that is, comets in orbits closer to the Sun which are pumped up to replace the lost comets^{19,27}.

Cometary capture is a highly improbable process because a three-body gravitational interaction is required to dissipate the excess energy of the hyperbolic orbit. The possibility of capture has been shown^{28,29} to be proportional to V_{∞}^{-7} , where V_{∞} is the hyperbolic-orbit excess velocity. Capture is possible at encounter velocities $\leq 1 \text{ km s}^{-1}$ but is highly unlikely at the Sun's velocity of 16.5 km s^{-1} relative to the local standard of rest.

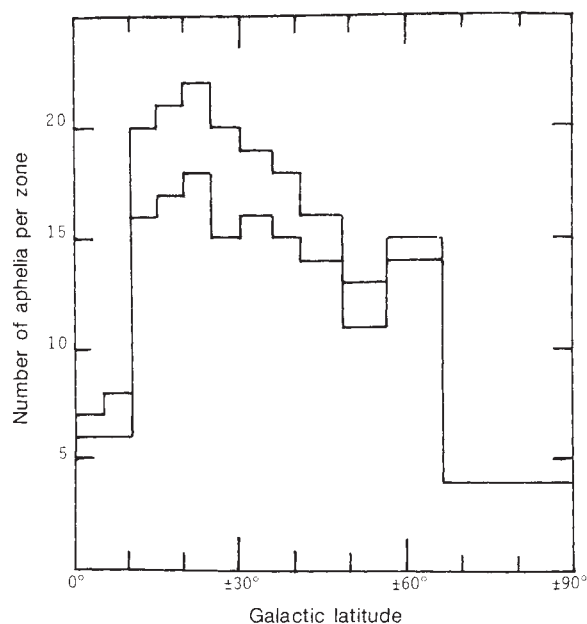


FIG. 2 Distribution in galactic latitude of aphelion directions for 152 long-period comets with orbital periods $> 10^4 \text{ yr}$ (ref. 22), plotted in bins of equal area on the sky. Comets are deficient at the galactic equator and poles, in agreement with the expectation from galactic tides acting on the Oort cloud. The lower curve is the raw data; the upper curve includes a correction for the lack of Southern Hemisphere observers.

Thus attention has turned to the possibility that comets lost from the Oort cloud may be replenished from an even more populous inner reservoir^{19,27}. Monte Carlo simulations have shown^{30,31} that the dynamical evolution of icy planetesimals formed in the Uranus–Neptune zone would naturally lead to a massive inner Oort cloud. The icy planetesimals would be perturbed by the evolving protoplanets into eccentric long-period orbits with semi-major axes of several times 10^3 AU. Galactic and stellar perturbations at aphelion would then be sufficient to raise the perihelia of such orbits to heliocentric distances beyond the orbit of Neptune, detaching the comets from planetary influences. In this manner, a massive inner Oort cloud reservoir, with a population between five and ten times that of the outer cloud, could be formed. As comets are stripped away from the outer cloud by close stellar and GMC encounters, the same perturbations will pump up comets from the inner cloud to replace them.

An example of one of the Monte Carlo simulations is shown in Fig. 3 for four different times in the Solar System's history³¹. Here the location of hypothetical comets in the Oort cloud are projected on a plane perpendicular to the galactic plane. Initially the comets are confined to the ecliptic plane (which is inclined $\sim 60^\circ$ to the galactic plane). As perturbations drive the orbits outward, their inclinations increase slowly. At the end of 4.5×10^9 yr the cloud is essentially randomized beyond $\sim 10^4$ AU from the Sun, but within that distance, the cloud remains flattened toward the ecliptic plane.

In this view, most of the comets originate as planetesimals in the Uranus–Neptune zone, although a small fraction come from both the Jupiter and Saturn regions of the solar nebula. Approximately 40–80% of the original comets ejected to the Oort cloud (both inner and outer) have been lost over the history of the Solar System^{27,31}. Comets are lost by direct ejection during close stellar and GMC passages, from perturbation back into the planetary region where the influences of Jupiter and Saturn will eject them hyperbolically or where they will be physically destroyed, and because of the slow growth of some orbits to distances greater than the Sun's sphere of influence, $\sim 2 \times 10^5$ AU.

Some researchers^{32–34} have suggested mechanisms for cometary formation *in situ*, in the more distant reaches of the protosolar nebula, as a means of overcoming the dynamical inefficiencies associated with populating the Oort cloud. But the problems of cometary accretion at such large solar distances have not been worked in detail, and such mechanisms remain somewhat speculative.

The differences between the inner and outer Oort clouds are matters of definition alone. There is one continuous cometary cloud surrounding the planetary system. The differences between the inner and outer clouds are defined by the different dynamical regimes: the outer Oort cloud is continuously stirred by distant stellar perturbations and the galactic tide to provide the steady-state flux of long-period comets into the planetary region. In comparison, the inner Oort cloud is dynamically inactive except in the presence of large perturbations such as penetrating stellar passages and GMCs.

Population and mass

Estimates of the population of the Oort cloud are obtained by comparing the flux calculated in Monte Carlo simulations with the observed number of long-period comets passing through the planetary region. The former is subject to statistical error as well as incomplete modelling of perturbations; the latter is a function of numerous observational selection effects. The best estimates so far are that the outer dynamically active Oort cloud has a population of $\sim 0.4\text{--}1.3 \times 10^{12}$ comets (brighter than absolute magnitude $H_{10} = 11$), as stated above, and the inner cloud has a population between 5 and 10 times that, or $2.0\text{--}13.0 \times 10^{12}$ comets. In my opinion, a best guess is a population of 1.0×10^{12} comets in the outer dynamically active Oort cloud, and 6×10^{12} comets in the unseen inner cloud.

The population estimates depend on the assumption that the presently observed long-period comet flux through the region of the inner planets is equivalent to the long-term average flux. If the present flux is enhanced owing to a recent perturbation on the Oort cloud, then the population estimate of the cloud is too high, and vice versa. Fernandez³⁵ has pointed out, however, that the galactic-latitude dependence of the long-period comet aphelion directions would not be present if most comets seen were the result of a recent large perturbation on the cloud by a close stellar passage or a GMC. Thus, the Oort cloud population estimates above may only be lower limits.

The total mass of the Oort cloud is not well known because of the large uncertainty in the bulk density of cometary nuclei^{36–38} and in the nucleus size distribution^{39,40}. Estimates for the total cloud mass range from ~ 14 to 1,000 Earth masses ($M_\oplus$), the upper limit being greater than the total mass of the planetary system^{39,41,42}. A best guess for the total mass is $45\text{--}50 M_\oplus$, assuming a nucleus bulk density of 0.6 g cm^{-3} and a mean nucleus mass of $3.8 \times 10^{16} \text{ g}$. Over the history of the Solar System, the Oort cloud has lost between 40% and 80% of its population, so the original mass must have been a factor of two to five larger.

One recent estimate⁴² of the Oort cloud mass also included an estimate for the angular momentum of $5 \times 10^{52}\text{--}2 \times 10^{53} \text{ g cm}^2 \text{ s}^{-1}$, far more than that of the entire planetary system. That estimate is, however, probably high. The total population for the inner and outer clouds was assumed to be a factor of 15 times greater than that given above, based only on some early speculations as to what the population might be¹⁹. Also, the conclusion that this high angular momentum in the Oort cloud means that the protoplanets could not have ejected so much material without spiralling in towards the Sun is incorrect; much of the present angular momentum in the cloud comes from the action of external perturbers over the history of the Solar System. The initial angular momentum must have been somewhat less.

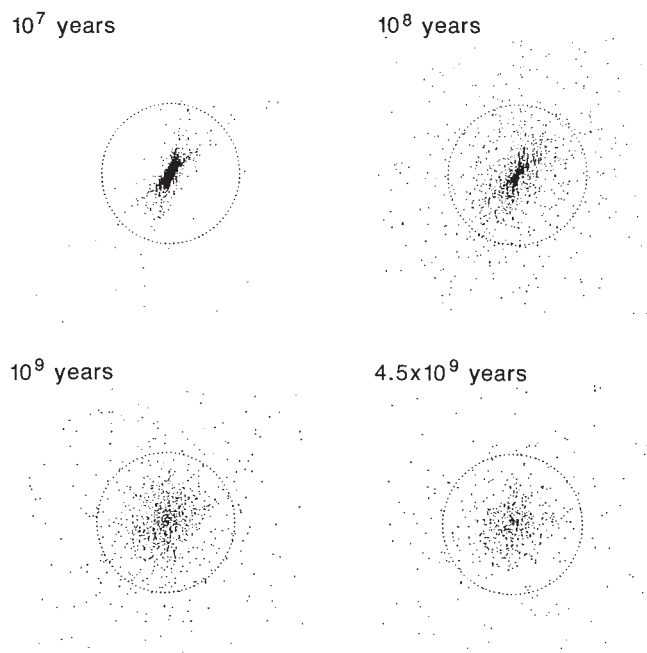


FIG. 3 Dynamical evolution of a hypothetical cloud of comets ejected out of the Uranus–Neptune zone, at several times during the history of the Solar System, under a combination of galactic, stellar and planetary perturbations (projected on a plane perpendicular to the galactic plane). The dotted circle is at a radius of 2×10^4 AU, the boundary between the inner and outer Oort clouds. From ref. 31.

Cometary flux variations

Monte Carlo studies have shown that variations in the expected flux from the outer Oort cloud by a factor of two to three are common, and variations by a factor of ten may occur occasionally^{23,43}. But more recent results⁴⁴ have indicated that the random fluctuations in the steady-state flux from the Oort cloud are probably only ~50%, possibly even less. The larger variations found in the earlier studies tended to reflect the limited size of the statistical samples, rather than true variations in the flux. An example of the expected flux into the planetary region from one of the newer simulations by Heisler⁴⁴ is shown in Fig. 4.

More extreme variations in the flux are possible if a star passes directly through the cloud, in particular through the inner Oort cloud¹⁹. A detailed Monte Carlo simulation⁴⁵ showed that a star passing at 3×10^3 AU from the Sun would produce a 2–3-Myr shower of $\sim 0.5 \times 10^9$ comets in Earth-crossing orbits, raising the expected impact rate by a factor of 300 or more. Such events are, however, rare and occur perhaps once every 300–500 Myr. Nevertheless, comet showers are expected to dominate the cometary contribution to cratering on the terrestrial planets⁴⁰. Such showers have been suggested as a possible cause of biological extinction events on the Earth¹⁹.

The suggestion from examination of the Earth's fossil record that biological extinction events might repeat with an approximately 26-Myr period⁴⁶ led to several hypotheses that invoked frequent comet showers as the cause of the extinctions. These hypotheses involved: (1) a dwarf companion star to the Sun in a distant eccentric 26-Myr-period orbit with perihelion deep in the Oort cloud^{47,48}; (2) a tenth planet circulating in a highly inclined, precessing orbit at ~ 150 AU from the Sun at the inner edge of the inner Oort cloud⁴⁹; and (3) the Solar System's vertical motion above and below the galactic plane with a half-period of 32–33 Myr, with GMC encounters at galactic-plane crossings⁵⁰. The apparent coincidence between the times of galactic-plane crossings by the Solar System and terrestrial extinction boundaries had been pointed out previously⁵¹. The Sun's vertical motion was also suggested as the clock mechanism by other researchers, although with only speculation about the underlying physical mechanism leading to the extinctions⁵².

A variety of dynamical problems have been identified with each of these hypotheses, and no evidence in support of any of them has been found. The orbit of the hypothetical companion star is unstable with an expected lifetime of $< 10^9$ yr (refs 53, 54). The orbital period of the dwarf star would vary by $\sim 10\%$ per orbit and would not maintain a fixed 26-Myr period²⁷. The precessing tenth planet, allegedly located in a cometary ring beyond the outer planets, would produce comet showers broadly distributed in time and with too few comets to result in a high probability of terrestrial impacts^{27,55}. Because the scale-height of GMCs above and below the galactic plane is almost the same as the maximum oscillation of the Sun and planetary system, the Oort cloud is almost as likely to encounter a GMC away from the plane as it would close to the plane⁵⁶. Therefore no periodic signal would result from such encounters. The alleged periodicity in the fossil extinction record has itself been questioned on statistical grounds⁵⁷, and on a lack of well dated tie points in the geological record⁵⁸.

A criticism applicable to all comet-shower hypotheses is that the observed cratering records on the Earth and Moon do not show the expected numbers of craters that would result from intense periodic comet showers every 26–32 Myr (refs 40, 59). This situation might be remedied if the Oort cloud population were smaller than the present estimate by a factor of four (or more), but that would also mean that showers would contain fewer comets, and thus the expected number of impacts on the Earth would be reduced accordingly. In addition, there is little evidence of multiple impacts, such as multiple craters, iridium layers and tektite layers, at extinction boundaries^{59,60}. This is somewhat surprising because at least one random comet shower would have been expected in the past 500 Myr.

The Kuiper belt

An interesting recent development has been the addition of a third component to the Oort cloud⁶¹. New Monte Carlo simulations have shown that most of the short-period comets probably do not result from the dynamical evolution of long-period comets. The long-period comets tend to preserve their random inclinations as they evolve inward, as a result of planetary perturbations, to short-period orbits. But the observed distribution of inclinations for the short-period comets is largely confined to direct orbits with inclinations $< 30^\circ$. Comet Halley is a notable exception. Only five retrograde short-period comets are known to exist.

Duncan and colleagues showed that a more plausible source for the short-period comets is a flattened belt or ring of comets beyond Neptune. This belt, presumably a remnant of the original protoplanetary disk of planetesimals in the solar nebula^{62–64}, is estimated to be ~ 300 times more efficient dynamically for producing short-period comets than direct evolution of long-period comets from the outer Oort cloud^{65–67}. The number of comets in the Kuiper belt, as Duncan and colleagues call it, is of the order of 10^8 – 10^{10} , with a total mass of perhaps 0.01 – $1.0 M_\oplus$.

Physical processing of comets

It had generally been thought that the Oort cloud was a fairly benign storage location for the long-period comets. The typical spacing between comets is 15 AU in the outer cloud and ~ 1 AU in the inner cloud. The typical temperature is that of interstellar space, ≤ 10 K. It is now recognized, however, that a variety of processes act on the cometary nuclei (in particular on their surface layers) over the history of the Solar System⁶⁸. These include: (1) irradiation, sputtering and polymerization by galactic cosmic rays⁶⁹; (2) heating by passing stars and nearby supernovae⁷⁰; (3) 'gardening' by cometary debris impacts⁷¹; and (4) the accretion of interstellar dust and gas and accompanying erosion by hypervelocity dust impacts^{72,73}.

The combined effects of these competing processes are not well understood. They may lead to the development of a permanent non-volatile crust over the nucleus surface before it ever enters the planetary system, or they may simply result in a heterogeneous collection of unbonded, highly processed materials on the well gardened nucleus surface.

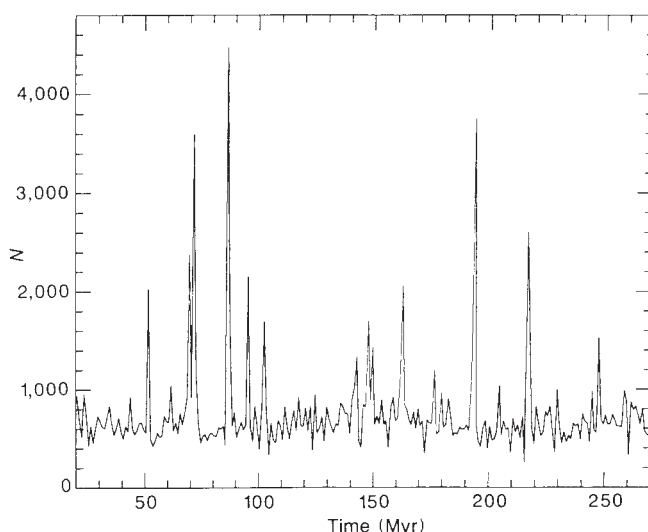


FIG. 4 Number of new long-period comets from the Oort cloud entering the planetary region (distance < 10 AU) as a function of time, based on a Monte Carlo simulation that included random passing stars and galactic perturbations. The large spikes are comet showers due to random stars penetrating the Oort cloud, but the small-scale variations still reflect the statistics of the sample size and not true variations in the flux. From ref. 44.

Extra-solar comet clouds and interstellar comets

Presumably the processes that led to the formation of planetesimals and their ejection to form an Oort cloud around our planetary system can also occur around other forming stars. Because the expected dimensions of Oort clouds are so large, it may be possible actually to detect and resolve such comet clouds. One method would be to look for thermal radiation from dust created by collisions and sputtering in the cloud. Such an analysis of sky images from the Infrared Astronomical Satellite has been performed for 29 nearby stars, looking for infrared excess in summed circular annuli around the primaries (S. A. Stern, personal communication). However, no certain detections have been reported. This is to be expected because radiation pressure and collisions with the interstellar medium should rapidly sweep fine dust from the clouds⁷³.

It is interesting to speculate on the fate of the many comets ejected to interstellar space in the formation and evolution of the Oort cloud. Dynamical ejection is the most common loss mechanism for comets in the cloud, either because of close stellar and GMC perturbations or as a result of perturbations by Jupiter during passage through the planetary system.

No comet on a clearly interstellar trajectory has been observed passing through the planetary system. This sets an upper limit on the local space density of interstellar comets⁷⁴ of $6 \times 10^{-4} M_{\odot} \text{pc}^{-3} \approx 4 \times 10^{12} \text{ comet pc}^{-3}$, using a mean nucleus mass of $3 \times 10^{17} \text{ g}$. For comparison, this is ~ 300 times less than the density of material in the solar neighbourhood ($\sim 0.185 M_{\odot} \text{pc}^{-3}$; ref. 75). It is about half the density for comets in the outer Oort cloud, assuming a population of 10^{12} comets in a sphere of radius 10^5 AU centred on the Sun. Thus, the limit is not very strict.

It is also possible to compare this limit with the expected space density of interstellar comets if it is assumed that all stars produce cometary clouds. From dynamical models^{76,77} it is estimated that between 3 and 50 times as many comets are ejected by the protoplanets as are placed in the Oort cloud (although it can be shown that the factor of 50 is extreme, as it assumes a very narrow range of semi-major axes for the Oort cloud). Another factor of two to three comes from the comets lost from the Oort cloud over the history of the Solar System. Thus, the Solar System has ejected $\sim 6 \times 10^{13} - 10^{15}$ comets to interstellar space. Taking a mean volume per star in the solar neighbourhood of 12 pc^3 (ref. 78), the predicted space density is $5 \times 10^{12} - 9 \times 10^{13} \text{ pc}^{-3}$, $\sim 1.3 - 23$ times the upper limit quoted above.

As half of all stars form in multiple systems, and that process may prevent the formation of a protoplanetary disk leading to cometesimals, the factor of 1.3 excess is probably not a problem.

A factor of 23 excess, however, is not consistent with the Oort cloud models and population estimates presented here, and thus represents a problem to be solved. As noted above, the factor of 50 ratio for ejected to bound comets is high by at least a factor of two (and probably four), but the remaining discrepancy still merits further study.

Conclusion

Our view of the Oort cloud has evolved considerably since it was first proposed in 1950. Much of that evolution has come in the past decade as a result of the availability of sufficient computing power to simulate the chaotic dynamical evolution of large numbers of hypothetical comets under a combination of perturbations. Another important factor has been a number of speculative hypotheses, often involving catastrophic events, which have not proven to be correct but which have nonetheless motivated researchers to improve their modelling and understanding of the dynamics of comets in the cloud. In addition, if comet clouds are a natural result of star and planet formation, we are approaching a time when such clouds may be detectable. The Hubble Space Telescope, the Infrared Space Observatory and other planned orbiting telescopes may soon provide the first conclusive evidence of Oort clouds around other stars, certainly one of the most exciting discoveries that we can anticipate. The existence of Oort clouds would strongly imply the existence of large planets, necessary to eject the protocomets to Oort cloud distances. The Hubble Space Telescope will also search for comets in the outer reaches of our own Solar System, in the hypothesized Kuiper belt beyond Neptune.

But our understanding of the Oort cloud is still far from complete. The observed distributions of orbital elements in galactic coordinates do not precisely match the expected distributions from Monte Carlo simulation of the Oort cloud. The average aphelion distance of dynamically new comets from the cloud is somewhat less than is expected from studies of the combined influences of random stars, GMCs and the galactic tide. Perhaps most important are the new studies of a possible comet belt beyond the orbit of Neptune, and its possible implications for the origin of the short-period comets.

The Oort cloud has 'grown' in number of comets, total mass and complexity over the past 10 years. But the central concept of a roughly spherical comet cloud surrounding the planetary system and stretching halfway to the nearest stars has remained remarkably intact. It will be interesting to see what changes to the current view are brought about by another decade of study.

P. Weissman is at the Earth and Space Sciences Division, Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91109, USA.

- Oort, J. H. *Bull. astr. Inst. Neth.* **11**, 91-110 (1950).
- Marsden, B. G., Sekanina, Z. & Everhart, E. *Astr. J.* **83**, 64-71 (1978).
- van Woerkom, A. F. F. *Bull. astr. Inst. Neth.* **10**, 445-472 (1948).
- Weissman, P. R. in *Dynamics of the Solar System* (ed. Duncombe, R. L.) 277-282 (Reidel, Dordrecht, 1979).
- Marsden, B. G., Sekanina, Z. & Yeomans, D. K. *Astr. J.* **78**, 211-225 (1973).
- Öpik, E. *Proc. Am. Acad. Sci.* **67**, 169-183 (1932).
- Halley, E. *A Synopsis of the Astronomy of Comets* (London, 1705).
- Weissman, P. R. in *Comets*, (ed. Wilkening, L. L.) 637-658 (Univ. of Arizona Press, 1982).
- Fernandez, J. A. *Astr. J.* **87**, 1318-1332 (1982).
- Everhart, E. *Astr. J.* **72**, 1002-1011 (1967).
- Weissman, P. R. in *Dynamics of Comets: Their Origin and Evolution*, (eds Carusi, A. & Valsecchi, G. B.) 87-96 (Reidel, Dordrecht, 1985).
- Bierman, L. in *Astronomical Papers Dedicated to Bengt Stromgren* (eds Reiz, A. & Anderson, T.) 327-335 (Copenhagen Observatory, 1978).
- Clube, S. V. M. & Napier, W. M. *Q. J. R. astr. Soc.* **23**, 45-66 (1982).
- Byl, J. *Moon Planets* **29**, 121-137 (1983).
- Harrington, R. S. *Icarus* **61**, 60-62 (1985).
- Heisler, J. & Tremaine, S. *Icarus* **65**, 13-26 (1986).
- Antonov, V. A. & Latyshev, I. N. in *The Motion, Evolution of Orbits and Origin of Comets* (eds Chebotarev, G. A., Kazimirchak-Polonskaya, E. I. & Marsden, B. G.) 341-335 (Reidel, Dordrecht, 1972).
- Smoluchowski, R. & Torbett, M. *Nature* **311**, 38-39 (1984).
- Hills, J. G. *Astr. J.* **86**, 1730-1740 (1981).
- Nezhinskij, E. M. in *The Motion, Evolution of Orbits, and Origin of Comets* (eds Chebotarev, G. A., Kazimirchak-Polonskaya, E. I. & Marsden, B. G.) 335-340 (Reidel, Dordrecht, 1972).
- Weissman, P. R. *Nature* **228**, 242-243 (1980).
- Delsemme, A. H. *Astr. Astrophys.* **187**, 913-916 (1987).
- Heisler, J., Tremaine, S. & Alcock, C. *Icarus* **70**, 269-288 (1987).
- Hut, P. & Tremaine, S. *Astr. J.* **90**, 1548-1557 (1985).
- Bailey, M. E. *Mon. Not. R. astr. Soc.* **218**, 1-30 (1986).
- Clube, S. V. M. & Napier, W. M. *Mon. Not. R. astr. Soc.* **208**, 575-588 (1984).
- Weissman, P. R. in *The Galaxy and the Solar System* (eds Smoluchowski, R., Bahcall, J. N. & Matthews, M. S.) 204-237 (Univ. of Arizona Press, 1986).
- Valtonen, M. J. *Observatory* **103**, 1-4 (1983).
- Valtonen, M. J. & Innanen, K. A. *Astrophys. J.* **255**, 307-315 (1982).
- Shoemaker, E. M. & Wolfe, R. F. *Lunar Planet Sci. Conf.* **XV** 780-781 (abstr.) (1984).
- Duncan, M., Quinn, T. & Tremaine, S. *Astr. J.* **94**, 1330-1338 (1987).
- Cameron, A. G. W. *Icarus* **18**, 407-450 (1973).
- Biermann, L. & Michel, K. W. *Moon Planets* **18**, 447-464 (1978).
- Bailey, M. E. *Icarus* **69**, 70-82 (1987).
- Fernandez, J. A. in *Comets in the Post-Halley Era* (eds Newburn, R. L., Rahe, J. & Neugebauer, M. M.) (Kluwer, Amsterdam, in the press).
- Ridman, H. in *The Comet Nucleus Sample Return Mission, European Space Agency, Rep. SP-249*, 195-205 (1986).
- Sagdeev, R. Z., Elyasberg, P. E. & Moroz, V. I. *Nature* **331**, 240-242 (1987).
- Peale, S. *Icarus* **84**, 36-49 (1989).
- Bailey, M. E. & Stagg, C. R. *Mon. Not. R. astr. Soc.* **235**, 1-35 (1988).
- Weissman, P. R. in *Global Catastrophes in Earth History, Geol. Soc. Am., Spec. Pap.* 247 (in the press).
- Weissman, P. R. *Bull. Am. astr. Soc.* **18**, 799 (abstr.) (1986).
- Marochnik, L. S., Mukhin, L. M. & Sagdeev, R. Z. *Science* **242**, 547-550 (1988).
- Fernandez, J. A. & Ip, W.-H. *Icarus* **71**, 46-56 (1987).
- Heisler, J. *Icarus* (submitted).

45. Hut, P. & Weissman, P. R. *Bull. Am. astr. Soc.* **17**, 690 (abstr.) (1985).
46. Raup, D. M. & Sepkoski, J. H. *Proc. natn. Acad. Sci. U.S.A.* **81**, 801-805 (1984).
47. Whitmire, D. P. & Jackson, A. A. *Nature* **308**, 713-715 (1984).
48. Davis, M., Hut, P. & Muller, R. A. *Nature* **308**, 715-717 (1984).
49. Whitmire, D. P. & Matese, J. J. *Nature* **313**, 36-38 (1985).
50. Rampino, M. R. & Stothers, R. B. *Nature* **308**, 709-712 (1984).
51. Innanen, K. A., Patrick, A. T. & Duley, W. W. *Astrophys. Space Sci.* **57**, 511-515 (1978).
52. Schwartz, R. D. & James, P. B. *Nature* **308**, 712-713 (1984).
53. Hills, J. G. *Nature* **311**, 636-638 (1984).
54. Torbett, M. V. & Smoluchowski, R. *Nature* **311**, 641-642 (1984).
55. Tremaine, S. in *The Galaxy and the Solar System* (eds Smoluchowski, R., Bahcall, J. N. & Matthews, M. S.) 409-416 (Univ. of Arizona Press, 1986).
56. Thaddeus, P. & Chanan, G. A. *Nature* **314**, 73-75 (1985).
57. Hoffman, A. *Nature* **315**, 659-662 (1985).
58. Shoemaker, E. M. & Wolfe, R. F. in *The Galaxy and the Solar System* (eds Smoluchowski, R., Bahcall, J. N. & Matthews, M. S.) 338-386 (Univ. of Arizona Press, 1986).
59. Weissman, P. R. *Nature* **314**, 517-518 (1985).
60. Hut, P. *et al.* *Nature* **329**, 118-126 (1987).
61. Duncan, M., Quinn, T. & Tremaine, S. *Astrophys. J.* **328**, L69-L73 (1988).
62. Kuiper, G. P. in *Astrophysics* (ed. Hynek, J. A.) 357-424 (McGraw Hill, New York, 1951).
63. Cameron, A. G. W. *Icarus* **1**, 13-69 (1962).
64. Cameron, A. G. W. in *The Origin of the Solar System* (ed. Dermott, S. F.) 49-75 (Wiley, New York, 1978).
65. Whipple, F. L. *Proc. natn. Acad. Sci. U.S.A.* **51**, 711-718 (1964).
66. Fernandez, J. A. *Mon. Not. R. astr. Soc.* **192**, 481-491 (1980).
67. Bailey, M. E. *Nature* **303**, 399-400 (1983).
68. Weissman, P. R. in *The Comet Nucleus Sample Return Mission, European Space Agency, Rep. SP-249* 15-25 (1986).
69. Johnson, R. E., Cooper, J. F., Lanzerotti, L. J. & Strazzula, G. *Astr. Astrophys.* **187**, 889-892 (1987).
70. Stern, S. A. & Shull, J. M. *Nature* **332**, 407-411 (1988).
71. Stern, S. A. *Icarus* **73**, 499-507 (1988).
72. Stern, S. A. *Icarus* **68**, 276-283 (1986).
73. Stern, S. A. *Icarus* (in the press).
74. Sekanina, Z. *Icarus* **27**, 123-133 (1976).
75. Bahcall, J. N. *Astrophys. J.* **276**, 169-181 (1984).
76. Fernandez, J. A. & Ip, W.-H. *Icarus* **47**, 470-479 (1981).
77. Safronov, V. S. *Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets, NASA, Rep. TT-F-677* (US Govt, 1972).
78. Allen, C. W. *Astrophysical Quantities* (Athlone, London, 1973).

ACKNOWLEDGEMENTS. I thank J. Heisler for discussions on Oort cloud dynamics and for the use of her unpublished results. I also thank M. Bailey, P. Campbell, J. Oort, A. Stern and D. Yeomans for reviews of an earlier version of this manuscript. This work was supported by the NASA Planetary Geology and Geophysics Program, and was performed at the Jet Propulsion Laboratory under contract with NASA.

ARTICLES

Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins

Hans-Michael Jantzen, Arie Admon, Stephen P. Bell & Robert Tjian

Howard Hughes Medical Institute, Department of Molecular and Cell Biology, Division of Biochemistry and Molecular Biology, 401 Barker Hall, University of California at Berkeley, Berkeley, California 94720, USA

The eukaryotic upstream binding factor (UBF), recognizes the ribosomal RNA gene promoter and activates transcription mediated by RNA polymerase I through cooperative interactions with the species-specific factor, SL1. Isolation of complementary DNA clones and sequence analysis reveals similarities between DNA binding domains of human UBF (hUBF) and high mobility group (HMG) protein 1. Expression, cellular localization and *in vitro* transcription studies establish that cloned hUBF encodes a nucleolar factor that binds specifically to the upstream control element and core of the rRNA gene promoter to activate transcription in a binding site-dependent manner.

REGULATION of transcription by eukaryotic RNA polymerases requires many accessory factors that direct specific protein-DNA recognition and protein-protein interactions¹⁻³. In contrast to RNA polymerase II and III, RNA polymerase I recognizes and transcribes from only one type of promoter which directs the transcription of the tandemly arrayed genes encoding the large rRNA precursors⁴. This unusual arrangement of a complete transcriptional apparatus dedicated to the expression of a single albeit critical cellular gene product, provides us with a unique opportunity to dissect its components and gain new insights into the biochemical function of sequence-specific regulatory proteins. In particular, a detailed analysis of RNA polymerase I transcription factors expressed from their cloned cDNAs should provide us with a better understanding of how

promoter-selective DNA-binding proteins interact productively with components of the transcriptional apparatus.

Analysis of the human rRNA promoter by *in vivo* transient transfection assays and *in vitro* transcription experiments has identified two important *cis* regulatory DNA elements, called the core and upstream control element (UCE)⁵⁻⁷. The core element extends from nucleotides -45 to +20 and mutations in this region essentially eliminate promoter activity. The UCE is located between -200 and -107 and stimulates transcription 10-100-fold in an orientation- and distance-dependent fashion. Despite extensive conservation of rRNA sequences from diverse organisms, rRNA gene promoters from different species show only very limited sequence similarity. More importantly, transcription extracts derived from one species will not transcribe a template derived from a heterologous source unless it is from a very closely related species (for example, human and monkeys)⁴.

Biochemical fractionation of the human transcription extract revealed at least two factors in addition to RNA polymerase I that are required for accurate and efficient transcription *in vitro*. One of these auxiliary factors (hSL1) has been extensively purified and is an essential component to reconstitute transcription⁸. In addition, hSL1 can confer human promoter specificity to a heterologous extract (for example, hSL1 can direct a mouse extract to transcribe the human rRNA template). It is of interest that hSL1 does not show any detectable sequence-specific DNA-binding activity on its own. But hSL1 can form a productive initiation complex on the DNA template in the presence of a second factor, hUBF, that binds specifically to the rRNA promoter. DNase I footprint experiments indicate that hUBF alone protects a region overlapping the UCE and interacts with similar DNA sequences contained in the core element. The addition of purified hSL1 leads to formation of a complex between the two

amino-acid sequence determined from six purified peptides. Two overlapping cDNA clones encoding hUBF were isolated from a HeLa cell cDNA library using oligonucleotide probes deduced from two peptide sequences (Fig. 1a). These cDNAs, which span 3,097 nucleotides, were sequenced, and an open reading frame encoding a polypeptide of 764 amino acids was found, starting with a ATG codon at nucleotide 148 and terminating with a stop codon at nucleotide 2,440. The open reading frame is preceded by an in-frame stop codon at nucleotide 94, suggesting that the complete coding sequence of hUBF had been cloned. Conceptual translation of this open reading frame predicts a highly charged protein (42% acidic and basic residues) containing all six peptide sequences obtained from purified hUBF. The calculated M_r is 89.4K, in good agreement with the size of the protein isolated from HeLa cells. The 3' end of the cloned cDNA consists of five adenines, but we did not find any candidate polyadenylation signal sequences at the appropriate distance from the 3' end of the cDNA (Fig. 1b).

Southern blot analysis of human genomic DNA with an hUBF cDNA probe detected a single species, indicating that hUBF is most likely encoded by a single gene (data not shown). The major hUBF messenger RNA in HeLa cells is of 3,200 nucleotides as determined by northern blot analysis, indicating that we had cloned nearly the entire major mRNA sequence (data not shown). In addition, a minor species of 4,500 nucleotides can be detected, which may be a consequence of alternatively spliced or processed mRNA. The human cDNA probe also efficiently hybridizes to two mRNA species of the same size in RNA from mouse leukaemia L1210 cells (data not shown). This result supports our finding that mouse cells express a transcription factor with the same sequence specificity and biochemical properties as human hUBF (H.-M.J., unpublished results; ref. 45)¹⁰.

Common structural domains

A homology search of databases with the hUBF amino-acid sequence did not reveal the presence of any known sequence-specific DNA-binding motifs. In addition, acidic amphipathic α -helices, or proline- or glutamine-rich domains involved in transcriptional activation of RNA polymerase II factors¹⁴⁻¹⁷, were not discernable in the primary sequence of hUBF. Unexpectedly, however, we found significant sequence similarity of hUBF to two members of a class of relatively abundant nuclear

proteins, HMG proteins 1 and 2 (HMG1 and HMG2)¹⁸. Although HMG proteins have not been shown to be sequence-specific DNA-binding factors, some display enhanced DNA-binding interactions with certain sequences such as AT-rich stretches¹⁹. The conserved sequences between hUBF and porcine HMG1 consist of three regions of 89, 83 and 86 amino acids (Fig. 2a). We have designated these conserved domains within the hUBF protein, HMG box-1, -2 and -3. An alignment of box 1 with HMG1 sequences from pig thymus (Fig. 2b) reveals a 40% conservation of primary structure, whereas a comparison with box 2 and 3 shows 35% and 42% similarity. Surprisingly, the HMG boxes in hUBF are not much more related to each other than each is to porcine HMG1.

The central region of the HMG boxes revealed some sequence similarity to part of the POU-specific domain found N-terminal to a homeobox in several factors involved in transcription and developmental regulation (data not shown)^{20,21}. The POU-specific domain is probably involved in efficient DNA binding, but its exact role remains unclear²². We also found some weak homology of hUBF to the yeast SPT2 gene product, a repressor of Ty element-controlled gene expression²³.

Another striking feature of the hUBF is the primary structure of its C terminus. Like HMG1 and 2, hUBF has a very acidic C terminus, but it is over twice the length of the HMG acidic tails (Fig. 2a). Of the terminal 89 amino acids, 64% are acidic, including two uninterrupted stretches of polyglutamic/aspartic acid residues of 21 and 18 amino acids, respectively. The remaining residues in the acidic C-terminus tail are exclusively serines (25%), asparagines (4%) and glycines (7%).

Functional identity of expressed UBF and hUBF

To study the structural basis of DNA binding, transcriptional activation and protein-protein interactions, wild-type and mutant versions of hUBF need to be expressed and isolated for biochemical studies. For some of these studies we used a vaccinia virus expression system which allows efficient expression of the protein. About 100 μ g vaccinia virus-derived hUBF (vUBF) was purified from 1 l infected HeLa cells to apparent homogeneity using a combination of conventional and DNA affinity chromatography. The purified protein migrated on SDS-PAGE gels as the 97K species of hUBF (Fig. 3a, lanes 1, 2). This indicates that the cloned cDNA encodes a full-length form of the protein purified from HeLa cells. Whether the 94K form

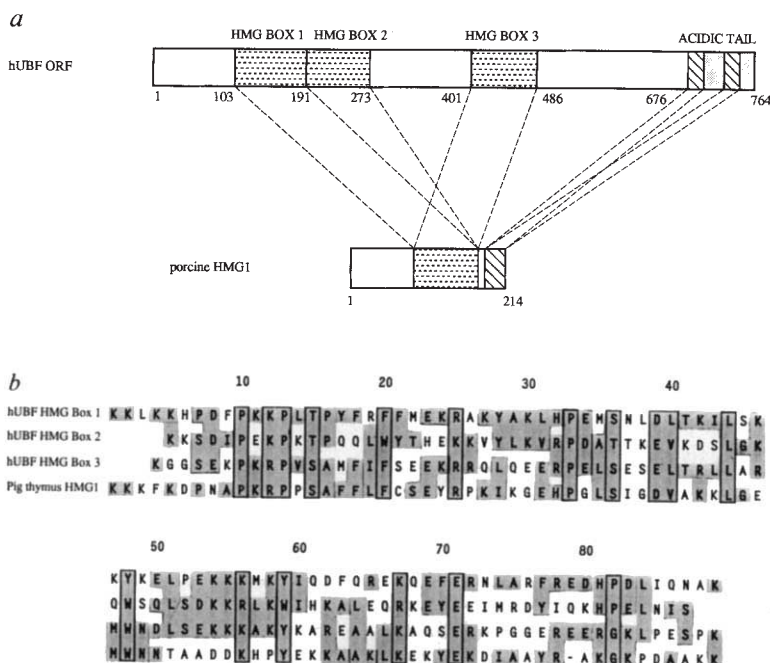
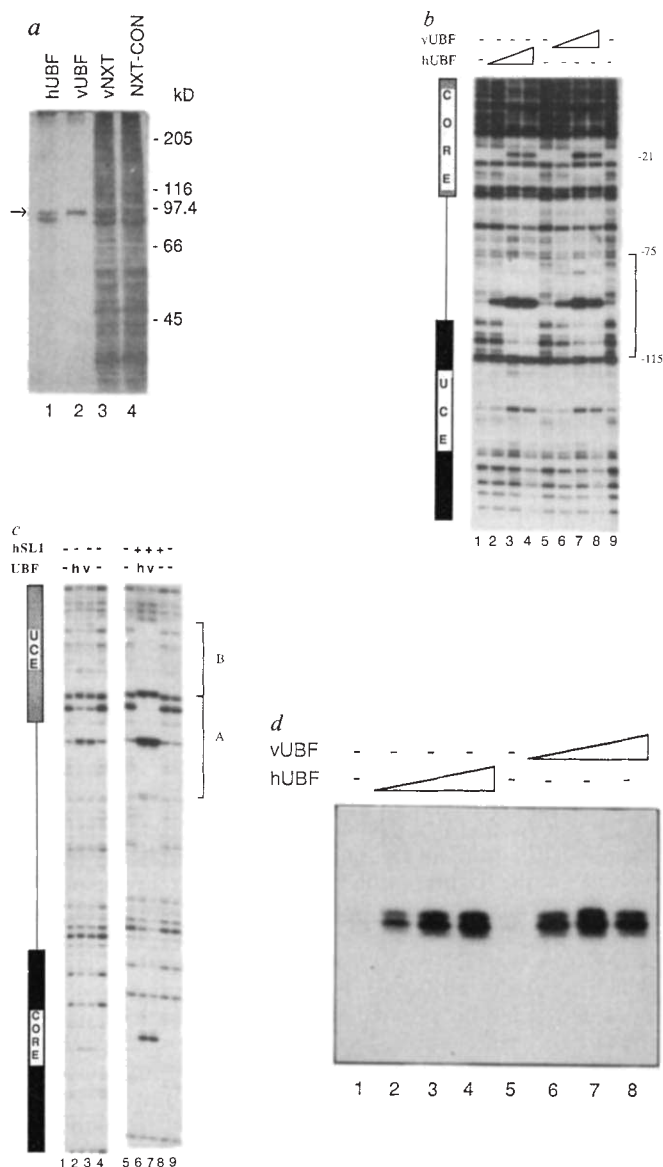


FIG. 2 Amino-acid sequence comparison of hUBF and HMG1. a, Comparison of the hUBF open reading frame with porcine HMG1. The stippled boxes represent amino-acid sequences similar to the putative DNA-binding domain of HMG1²⁸, whereas the cross-hatched boxes indicate polyacidic stretches in both proteins. The shaded segments in the C terminus of hUBF signify sequences rich in acidic amino acids. The numbering indicates amino-acid positions. b, Amino-acid sequence alignment of hUBF with porcine HMG1 (single-letter code). The first line shows hUBF amino acids 103-191 (HMG box 1), the second line, hUBF amino acids 191-273 (HMG box 2), the third line, hUBF amino acids 401-486 (HMG box 3), and the lowest line, porcine HMG1 amino acids 85-172 (allowing for one gap between positions 162 and 163). The numbering indicates positions within the HMG box. Amino acids identical or conserved in all four sequences are boxed, and those identical or conserved in at least two out of four positions are shaded. Conserved amino-acid substitutions are grouped as follows: (K,R), (E,D), (Q,N), (S,T), (F,Y,W,H), (L,I,V,A), (P,G) and (M,C).

FIG. 3 DNA binding and transcriptional properties of vUBF expressed from recombinant vaccinia virus. **a**, Purification of vUBF expressed from recombinant vaccinia virus. SDS-PAGE of 150 ng affinity-purified hUBF from HeLa cells (lane 1), 100 ng affinity-purified vUBF expressed from vaccinia virus (lane 2), 25 μ g nuclear extract from HeLa cells infected with recombinant vaccinia virus (lane 3), 25 μ g nuclear extract from HeLa cells infected with wild-type vaccinia virus (lane 4). Proteins were visualized by silver-staining. The positions of the M_r size marker proteins are shown on the right. **b**, Binding of recombinant vUBF to the rRNA gene promoter. DNase footprinting reactions were performed on the noncoding strand of the human rRNA promoter; the positions of the promoter elements are indicated on the left. Lanes 1, 5 and 9—no protein; lanes 2, 3 and 4—0.3, 1.2 and 3.6 ng affinity-purified hUBF from HeLa cells, respectively; lanes 6, 7 and 8—similar amounts of affinity-purified vUBF expressed from recombinant vaccinia virus. The bracket on the right shows the protected region; asterisk indicates enhanced cleavage in the core element. **c**, Interaction of SL1 with recombinant vUBF. Footprint reactions were performed on the coding strand of the human rRNA promoter with 4 ng affinity-purified HeLa hUBF (lane 2) or vUBF expressed from recombinant vaccinia virus alone (lane 3) and with the addition of 0.8 μ g hSL1 (lanes 6 and 7). Lane 8, hSL1 alone; lanes 1, 4, 5 and 9, from reactions without proteins. The position of the promoter elements is indicated on the left, whereas the brackets on the right show regions protected in the presence of UBF alone (A) or upon addition of both UBF and hSL1 (B). **d**, Transcriptional activation by vUBF expressed from vaccinia virus. An S1 nuclease analysis of *in vitro* transcription reactions using the wild-type human promoter is shown. All reactions contain 2 μ g partially purified human RNA polymerase I and 0.6 μ g hSL1. Lanes 2, 3 and 4 contain in addition 0.04, 0.12 and 0.5 ng affinity-purified hUBF from HeLa cells, respectively, whereas lanes 6, 7 and 8 contain similar amounts of affinity-purified vUBF expressed from recombinant vaccinia virus.

METHODS. For the construction of the UBF vaccinia expression vector, plasmid pVUBF2, both hUBF cDNAs were fused at the common *Bgl*III site in the overlapping region (position 1,616, see Fig. 1b), subcloned in pBlue-scriptSK⁺ and the sequence at the first ATG converted into a *Nde*I site by site-directed mutagenesis. From the resulting construct, pSUBF1Nde, the *Nde*–*Hind*III fragment (positions 147–2,839) containing the complete open reading frame was cloned into the *Xba*I site of the vaccinia virus expression vector pAbt4587 (kindly provided by Applied Biotechnology) after filling in the ends with Klenow. Using this vector, pVUBF2, recombinant virus was generated essentially as described⁴⁰. HeLa cells at a density of 5×10^8 cells l^{-1} were infected at a multiplicity of infection of 1.0, with either the recombinant virus or a New York City Board of Health (NYCBH) strain of vaccinia virus (kindly provided by Applied Biotechnology). Forty-eight hours after the beginning of infection, a nuclear extract was prepared as described⁴¹, except that the buffers contained additional 10 μ g ml^{-1} leupeptin, 10 μ g ml^{-1} pepstatin A, 2 μ g ml^{-1} aprotinin and 2 mM benzamide. UBF expressed from vaccinia virus (vUBF) was purified to homogeneity from these extracts using Q Sepharose Fast Flow (Pharmacia-LKB, step elution with 0.5 M KCl after a 0.2 M KCl wash) and one or two rounds of sequence-specific DNA affinity chromatography (see Fig. 1). Material prepared with the same procedure but from HeLa cells infected with wild-type vaccinia virus (NYCBH strain) contained at least 100-fold less UBF footprinting activity (H.-M.J., unpublished observations). SDS-PAGE and silver staining were



performed using standard procedures. Footprinting and *in vitro* transcription reactions were conducted as previously described^{9,10}, except that UBF was preincubated with the transcription template for 10 min on ice before adding the other components.

arises from differential splicing, an alternative translational start site, post-translational modification or proteolysis, remains to be investigated.

The DNA-binding specificity of vUBF was investigated in a DNase footprinting assay using the human rRNA gene promoter as a template (Fig. 3b). Essentially identical DNase cleavage patterns were observed with increasing amounts of hUBF from HeLa cells (lanes 2–4) and vUBF expressed from recombinant vaccinia virus (lanes 6–8). Both proteins protect the region between –75 and –115 overlapping the UCE. In addition, both proteins interact with the core element of the promoter, as evidenced by an enhanced DNase cleavage at position –21. A characteristic property of hUBF is its cooperative interaction with transcription factor hSL1 to activate transcription by RNA polymerase I (refs 9, 10). This protein–protein interaction can be detected as an enhancement and extension of the hUBF footprint into the critical region of the UCE (region B of the footprint), when hSL1 is present in the DNA-binding reaction (Fig. 3c, lane 6 versus lane 2). Indeed, vUBF forms an identical complex with hSL1, indicating that this important property of

protein–protein interaction is retained (Fig. 3b, lane 7 versus lane 3).

The transcriptional activation properties of cloned hUBF were analysed in a reconstituted *in vitro* reaction containing partially purified RNA polymerase I and hSL1 from HeLa cells and various amounts of hUBF¹⁰. In the absence of hUBF, a very low basal level of transcription is detected by S1-nuclease analysis of the RNA products (Fig. 3d, lanes 1 and 5). Increasing amounts of either hUBF (lanes 2–4) or vUBF (lanes 6–8) leads to strong activation of transcription (more than 10-fold). We conclude that hUBF expressed in the vaccinia system is a fully active recombinant protein useful for further functional studies.

DNA binding by hUBF involves the HMG box

To delineate the functional domains of hUBF, truncated forms of the protein were generated by *in vitro* translation. The DNA-binding properties of these deletion mutants were assayed by an analytical DNA-affinity-column assay, previously applied to the characterization of Fos–Jun complexes²⁴. Mutants of hUBF with a sequence-specific DNA-binding domain bind to a DNA

affinity column bearing UBF-binding sites in the presence of nonspecific competitor and elute at higher salt concentrations, whereas mutants lacking these domains flow through at low salt concentrations.

Protein synthesized *in vitro* from a wild-type hUBF template migrated on SDS gels with the same mobility as HeLa hUBF (Fig. 4, wt panel, IN lane). As is commonly observed, some full-length hUBF did not bind to the column and appeared in the flow-through or wash fractions. But more than 50% of the input hUBF bound to the column and could be eluted at high salt concentration, indicative of specific binding to the hUBF recognition sequence. When DNA binding was performed with a deletion mutant containing the N-terminal 489 amino acids ($\Delta C489$) which includes all three HMG boxes but not the acidic tail (see Fig. 4, diagram), essentially wild-type levels of DNA-binding activity were observed. A deletion mutant including HMG box 1 ($\Delta C204$), but lacking HMG box 3, most of box 2 and the C terminus, also showed efficient DNA binding relative to the full-length protein. By contrast, a mutant ($\Delta C90$) containing only the N-terminal 90 amino acids, which lacks all HMG boxes, failed to bind the column, and all the truncated protein could be detected in flow-through or low-salt wash fractions. Taken together, these data indicate that the acidic C-terminal tail is unnecessary for efficient DNA binding. In addition, at least part of HMG box 1 is necessary and sufficient for DNA binding of hUBF in this assay, supporting the hypothesis that HMG boxes constitute a new class of DNA-binding domains (see discussion).

hUBF is a nucleolar transcription factor

If UBF is a specific RNA polymerase I transcription factor it should be localized to the nucleolus, in contrast to factors involved in the transcription of mRNA that are generally excluded from the nucleolus. To address this question, we have raised rabbit polyclonal antibodies against a synthetic peptide derived from the cDNA sequence in a region just N-terminal to the hyperacidic tail (hUBF PepC; see Fig. 5, methods). The antiserum was affinity-purified on a matrix to which the antigen peptide had been coupled. In a western blot experiment, the anti-UBF PepC antibodies specifically recognize the characteristic hUBF doublet in partially purified or homogeneous hUBF preparations from HeLa cells (Fig. 5a, lanes 3, 4). The antibodies also detect the single species of vUBF protein expressed from recombinant vaccinia virus after affinity purification (Fig. 5a, lane 5) or in crude nuclear extracts from infected HeLa cells

(Fig. 5a, lane 6). By contrast, preimmune antibodies do not recognize any specific proteins under these conditions (Fig. 5a, lanes 1, 2). We have previously reported the existence of an hUBF-like protein in mouse cells¹⁰ and have detected a mouse mRNA with the hUBF probe (data not shown). Here we find that antibodies raised against PepC of hUBF efficiently cross-reacts with mUBF (also a doublet of 97K and 94K, see Fig. 5a) partially purified from mouse cells. These results confirm our prediction that the UBF proteins from mouse and human are highly conserved (H.-M.J., unpublished observations; ref. 45).

Having established the specificity of anti-UBF PepC, we used these antibodies in immunofluorescence microscopy experiments with HeLa cells. Antibody staining of these cells indicates that hUBF is predominantly localized in the nucleolus, the intracellular site of RNA polymerase I-mediated transcription (Fig. 5b). In addition, some weak, patchy fluorescence was detected in the nucleus, which could indicate a role for hUBF in the nucleus as well, or reflect a general background detected by these antibodies. As expected, the preimmune antibodies did not show any significant staining except for a faint background level in the nucleus and cytoplasm (Fig. 5d). By contrast, an antiserum raised against Sp1, a defined RNA polymerase II transcription factor²⁵, stains exclusively the nucleus, with no fluorescence in the nucleolus (Fig. 5c). These data strongly support our hypothesis that hUBF is a specific DNA-binding protein responsible for the transcriptional activation of the large rRNA precursors by RNA polymerase I in the nucleolus.

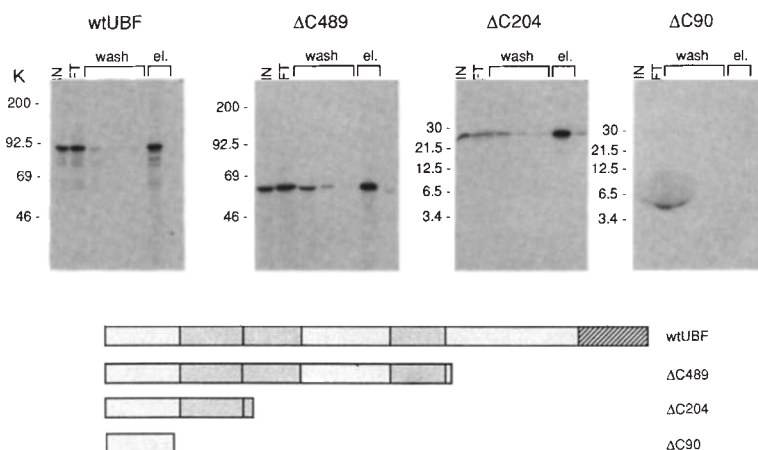
Discussion

Does the HMG box represent a novel class of DNA-binding domain?

Transcription factor hUBF reveals a striking similarity to the nuclear nonhistone proteins HMG1 and 2 which are thought to play a part in chromosome structure and gene activity¹⁸. HMG1 and 2 bind to AT-rich sequences *in vitro* and prefer single-stranded to double-stranded DNA¹⁹. Although HMG proteins are associated with the chromatin of only a subset of genes, their relative abundance and lack of sequence specificity compared with transcription factors seem to exclude a gene-specific regulatory role¹⁸. So it was surprising that the regions of homology between HMG1, HMG2 and hUBF coincide with the nonspecific DNA-binding domain of HMG1 and 2 (ref. 26). In HMG1 and 2 this domain has a high α -helical content¹⁸. Similarly, the HMG boxes of hUBF display a strong potential to form α -helices, as judged by Chou and Fasman analysis (H.-M.J., unpublished observations). Such α -helical domains

FIG. 4 Binding of hUBF deletion mutants translated *in vitro* to UBF binding site DNA columns. Wild-type and C-terminal deletions of UBF were transcribed and translated *in vitro* incorporating [³⁵S]methionine and passed over small DNA affinity columns containing the UBF recognition sequence in the presence of nonspecific competitor. The columns were washed at low salt concentration and specifically bound protein was eluted with high salt-containing buffer. Aliquots of the fractions were separated by SDS-PAGE and autoradiographed. Lanes of the gels are designated as follows: IN, input material; FT, flow-through; wash, low-salt wash fractions; el, high-salt eluate fractions. The numbers on the left indicate positions of M_r markers ($M_r \times 10^{-3}$). The structure of the C-terminal deletion mutants is shown below, the stippled boxes representing the HMG boxes, and the cross-hatched box, the acidic tail of hUBF.

METHODS. The *in vitro* transcription-translation vector pT7GUBF1 was constructed by ligating the *Nde*I-*Hind*III UBF cDNA fragment of pSUBF1Nde (see Fig. 3) after treating the ends with mung-bean nuclease into the *Stu*I-*Hinc*II sites of pT7 β Stu (a pGEM4 derivative containing the β -globin gene leader downstream of the T7 polymerase promoter, provided by R. Treisman⁴²). The vector was cut at various restriction sites and transcribed *in vitro* with T7 RNA polymerase. *In vitro* translation of the resulting RNAs and analytical DNA affinity chromatography was performed as described²⁴ with the following modifications: column buffer was TM⁺ (ref. 10), the loaded protein



material contained 1 μ g poly[d(AT)-d(AT)] and 0.05 μ g sonicated salmon sperm DNA as nonspecific competitor. Loading and washing buffer contained 0.1 M KCl, elution was performed at 0.6 M salt. Fractions from wtUBF and $\Delta C489$ were separated on 8% polyacrylamide gels, and $\Delta C204$ and $\Delta C90$ were separated on 15% gels.

have frequently been implicated in directing protein-DNA interactions in other DNA-binding proteins^{1,2}. The strongest evidence for a role of the HMG box in DNA binding comes from the preliminary domain mapping experiments using C-terminal deletion mutants of hUBF in an analytical DNA affinity chromatography assay. These results indicate that the presence of HMG box 1 is necessary and sufficient for DNA binding. We therefore suggest that the HMG boxes of hUBF represent a novel class of DNA-binding domain, and it is likely that other regulatory proteins with this structure will be found.

In contrast to the AT-rich sequences bound by HMG1 and 2, UBF from a variety of species binds specifically to sequences

that are generally, but not exclusively, GC-rich (H.-M.J., unpublished observations; ref. 45)^{10,11}. More importantly, the interaction of UBF with DNA is sequence-specific, whereas that of HMG1 and 2 is not. It is possible that the HMG domain of UBF has evolved from a primordial nonspecific DNA-binding structure. This is supported by the absence of a clear consensus DNA sequence for UBF binding sites¹¹⁻¹³ which could indicate that the DNA-binding domain of UBF is adapted to interact with diverse rRNA gene promoter sequences in different species. The sequence specificity of UBF is also significantly influenced by protein-protein interaction with SL1 (ref. 45).

Another difference between HMG1 and 2 and hUBF is that the latter contains three HMG boxes. It is tempting to speculate that each box is capable of binding to DNA on its own with weak affinity, but that together they allow high-affinity binding to adjacent sites on the template. Alternatively, one HMG box could be mostly responsible for nonspecific DNA binding, whereas others would direct sequence-specific contacts. Detailed mutational analysis and protein-DNA interaction studies are necessary to clarify these points.

The hyperacidic hUBF tail might interact with SL1. A second unusual structural feature of the hUBF protein is its hyperacidic C terminus, which includes two polyacidic stretches of 18 and 21 amino acids. Such hyperacidic regions have been previously identified in many proteins with diverse functions and cellular localization²⁷. Mammalian HMG1, for example, contains a C terminus of 30 acidic amino acids attached to a DNA-binding domain homologous to that of hUBF²⁸. In this case, the acidic structure binds the positively charged histones *in vitro* and has been proposed to interact with nucleosome cores contributing to some role in chromatin assembly-disassembly *in vivo*²⁹. HMG1 and 2 also stimulate transcription by RNA polymerase II and III *in vitro*^{30,31}, whereas another HMG1-like protein that is unrelated to HMG1 and 2, or the HMG boxes of hUBF, stimulates rat rRNA transcription *in vitro*³². However, the specificity of this activation and the contribution of acidic tails remains unclear. It is of interest that most cloned nucleolar proteins^{33,34} contain polyacidic stretches. These may allow interaction with positively charged surfaces of proteins such as histones and ribosomal proteins. By analogy, we propose that the hyperacidic tail of hUBF interacts with the essential RNA polymerase I transcription factor hSL1¹⁰. Chromatographic studies indicate that hSL1 is a highly positively charged protein as judged by its strong binding to negatively charged matrices⁸ (M. Learned and H.-M.J., unpublished observations). Therefore an ionic interphase between the hyperacidic tail of hUBF and hSL1 could mediate the interaction between these two factors. But other domains are probably also involved in conferring the specificity of this interaction. The hUBF sequence also contains possible phosphorylation sites for casein kinase II³⁵. Phosphorylation or other post-translational modifications of hUBF could potentially modulate the interaction with hSL1 and therefore be part of a regulatory mechanism.

Although acidic domains are involved in transcriptional activation by various RNA polymerase II factors, they are generally thought to consist of negatively charged amphipathic α -helices rather than long polyacidic stretches¹⁵. This indicates that the molecular mechanism by which the C terminus of hUBF functions to form a transcription complex could be significantly different from other acidic activators. Preliminary experiments indicate that hUBF cannot activate transcription by RNA polymerase II *in vitro* (S.P.B., unpublished results).

Protein-protein interactions between UBF and SL1 are a principal determinant of the observed species specificity of rRNA transcription by RNA polymerase I between frog and human cells¹². For example, xUBF from *Xenopus laevis* will not activate human RNA polymerase I, even though its DNA-binding properties are indistinguishable from those of hUBF. With the availability of the hUBF cDNA clone, we should now be able to delineate the structural domains of hUBF responsible

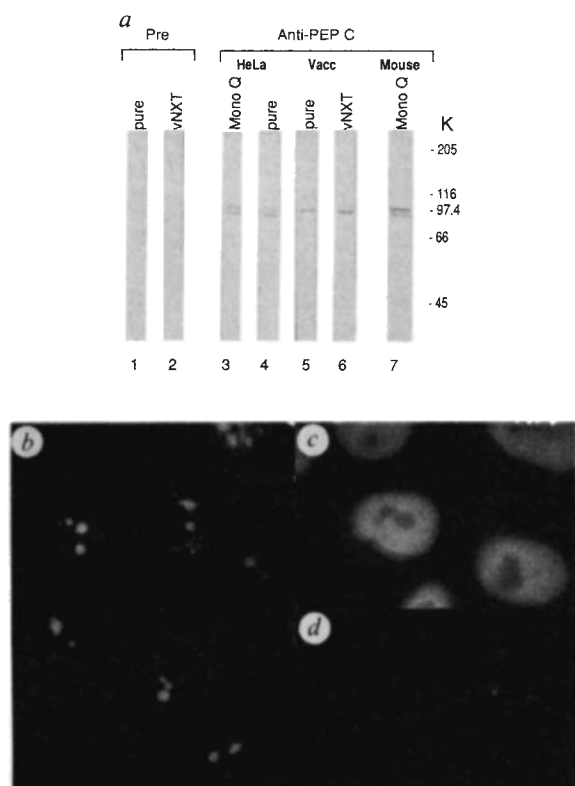


FIG. 5 Nucleolar localization of hUBF with anti-UBF peptide antibodies. *a*, Western blot. Various samples of pure and partially purified UBF from different sources were subjected to western blot analysis using affinity-purified antibodies against peptide C (lanes 3-7) or similarly treated pre-immune serum (lanes 1 and 2). Antibody dilution was 1:4,000. Lanes 1 and 5, 25 ng affinity purified UBF expressed from recombinant vaccinia virus; lanes 2 and 6, 25 μ g nuclear extract from HeLa cells infected with recombinant vaccinia virus; lane 3, 25 μ g partially purified HeLa UBF (mono Q fraction); lane 4, 20 ng of affinity-purified HeLa UBF; lane 6, 40 μ g partially purified mouse UBF (mono Q fraction). *b*, Immunofluorescence microscopy of HeLa cells with anti-UBF peptide antibodies (1:100 dilution). *c*, Immunofluorescence microscopy of HeLa cells with anti-Sp1 antiserum (1:500 dilution). *d*, Immunofluorescence microscopy of HeLa cells with preimmune serum (1:100 dilution). Positions of M_r size markers are shown on the right ($M_r \times 10^{-3}$).

METHODS. HeLa and mouse (L1210 cell) UBF were purified as described in Fig. 1; vaccinia UBF was prepared as in Fig. 3. Antibodies against UBF peptide C (PepC) were raised in rabbits and affinity-purified on Affigel 10 (Biorad) coupled with PepC using standard procedures⁴³. PepC (-CSNKRKSMTKLR-GPNPKSSR-) (N- to C-terminal; single-letter code, hUBF amino acids 650-668 with cysteine added to the N terminus) was synthesized by Milligen-Bioscience. Anti-Sp1 antiserum was a gift of S.P.Jackson. Gel electrophoresis and western blot analysis was as described⁴⁴ with alkaline phosphatase-coupled secondary antibodies. For immunofluorescence microscopy, HeLa cells were grown on coverslips, fixed for 5 min with 70% ethanol, 7 min with 100% ethanol, and 5 min with 70% ethanol, all at -20°C . Goat anti-rabbit antibody conjugated with biotin and fluorescein-avidin (both from Vector Laboratories) were used for antibody detection.

for mediating this species-specific interaction. The specificity of multiprotein complexes could be an important feature not only for transcription by RNA polymerase I, but also regulation of transcription initiation by RNA polymerase II and III.

How is hUBF localized to the nucleolus? The finding that the main site of hUBF action coincides with the nucleolar location of rRNA transcription strongly supports our hypothesis that hUBF is a genuine RNA polymerase I factor. But we cannot exclude the possibility that hUBF plays a part in the nucleolus in addition to being a transcription factor, for example in processing and transport of rRNA and preribosomal particles. Although the nucleolar localization of hUBF could be mainly due to its affinity for DNA-binding sites on the reiterated rRNA gene promoters, it is more likely that a distinct nucleolar localization signal in

the hUBF protein sequence directs it to the site of action, as occurs for the pX protein of human T cell leukaemia virus type I (HTLV-I)³⁶. The molecular cloning of hUBF should now allow the mapping of nucleolar localization signals by transfecting fusion constructs containing portions of hUBF linked to non-nucleolar marker proteins which can be detected with antibodies.

Note added in proof: Since submission of this manuscript we have detected a fourth HMG box in hUBF (amino acids 287–372) which is more diverged from HMG1 than the other HMG boxes. Additionally, sequence analysis of another DNA-binding protein, the yeast ARS binding protein ABF2, revealed two HMG boxes (J. Diffley and B. Stillman, personal communication). □

Received 29 December 1989; accepted 27 February 1990.

- Mitchell, P. J. & Tjian, R. *Science* **245**, 371–378 (1989).
- Johnson, P. F. & McKnight, S. L. *A. Rev. Biochem.* **58**, 799–839 (1989).
- Saltzman, A. G. & Weinman, R. *FASEB J.* **3**, 1723–1733 (1989).
- Sollner-Webb, B. & Tower, J. A. *Rev. Biochem.* **55**, 801–830 (1986).
- Smale, S. T. & Tjian, R. *Molec. cell. Biol.* **5**, 352–362 (1985).
- Haltiner, M. M., Smale, S. T. & Tjian, R. *Molec. cell. Biol.* **6**, 227–235 (1986).
- Jones, M. H., Learned, R. M. & Tjian, R. *Proc. natn. Acad. Sci. U.S.A.* **85**, 669–673 (1988).
- Learned, R. M., Cordes, S. & Tjian, R. *Molec. cell. Biol.* **5**, 1358–1369 (1985).
- Learned, R. M., Learned, T. K., Haltiner, M. M. & Tjian, R. *Cell* **45**, 847–857 (1986).
- Bell, S. P., Learned, R. M., Jantzen, H.-M. & Tjian, R. *Science* **241**, 1192–1197 (1988).
- Pikaard, C. S., McStay, B., Schults, M. C., Bell, S. P. & Reeder, R. H. *Genes Dev.* **3**, 1779–1788 (1989).
- Bell, S. P., Pikaard, C. S., Reeder, R. H. & Tjian, R. *Cell* **59**, 489–497 (1989).
- Dunaway, M. *Genes Dev.* **3**, 1768–1778 (1989).
- Courey, A. J., Holtzman, D. A., Jackson, S. P. & Tjian, R. *Cell* **59**, 827–836 (1989).
- Ptashe, M. *Nature* **335**, 683–689 (1988).
- Courey, A. J. & Tjian, R. *Cell* **55**, 887–898 (1988).
- Mermoud, N., O'Neill, E. A., Kelly, T. J. & Tjian, R. *Cell* **58**, 741–753 (1989).
- Einck, L. & Bustin, M. *Exp. cell Res.* **156**, 295–310 (1985).
- Wright, J. M. & Dixon, G. H. *Biochemistry* **27**, 576–581 (1988).
- Herr, W. *et al. Genes Dev.* **2**, 1513–1516 (1988).
- He, K. *et al. Nature* **340**, 35–42 (1989).
- Sturm, R. A. & Herr, W. *Nature* **336**, 601–604 (1988).
- Roeder, G. S., Beard, C., Smith, M. & Keranen, S. *Molec. cell. Biol.* **5**, 1543–1553 (1985).
- Turner, R. & Tjian, R. *Science* **243**, 1689–1694 (1989).
- Kadonaga, J. T., Jones, K. A. & Tjian, R. *Trends biochem. Sci.* **11**, 20–23 (1986).
- Carballo, M., Puigdomenech, P. & Palau, J. *EMBO J.* **2**, 1759–1764 (1983).

- Earnshaw, W. C. *J. Cell. Biol.* **105**, 1479–1482 (1987).
- Tsuda, K., Kikuchi, M., Mori, K., Waga, S. & Yoshida, M. *Biochemistry* **27**, 6159–6163 (1988).
- Bernues, J., Espel, E. & Querol, E. *Biochim. biophys. Acta* **866**, 242–251 (1986).
- Tremethick, D. J. & Molloy, P. L. *J. biol. Chem.* **261**, 6986–6992 (1986).
- Watt, F. & Molloy, P. L. *Nucleic Acids Res.* **16**, 1471–1486 (1988).
- Yang-Yen, H.-F. & Rothblum, L. I. *Molec. cell. Biol.* **8**, 3406–3414 (1988).
- Lapeyre, B., Bourbon, H. & Amalric, F. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1472–1476 (1987).
- Schmidt-Zachmann, M. S., Hugle-Dorr, B. & Franke, W. W. *EMBO J.* **6**, 1881–1890 (1987).
- Kuenzel, E. A., Mulligan, J. A., Sommercorn, J. & Krebs, E. G. *J. biol. Chem.* **262**, 9136–9140 (1987).
- Siomi, H. *et al. Cell* **55**, 197–209 (1988).
- Bohmann, D. *et al. Science* **233**, 1386–1392 (1987).
- Lathe, R. *J. molec. Biol.* **183**, 1–12 (1985).
- Kadonaga, J. T., Carner, K. R., Masiarz, F. R. & Tjian, R. *Cell* **51**, 1079–1090 (1987).
- Chakrabarti, S., Brechling, K. & Moss, B. *Molec. cell. Biol.* **5**, 3403–3409 (1985).
- Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. *Nucleic Acids Res.* **11**, 1475–1489 (1983).
- Norman, C., Runswick, M., Pollock, R. & Treisman, R. *Cell* **55**, 989–1003 (1988).
- Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1988).
- Jackson, S. P. & Tjian, R. *Cell* **55**, 125–133 (1988).
- Bell, S. P., Jantzen, H.-M. & Tjian, R. *Genes Dev.* (in the press).

ACKNOWLEDGEMENTS. We thank B. O'Neill and B. Johnson for help in generating recombinant vaccinia viruses, all members of the lab for helpful discussions and support, G. Gill and R. Turner for critically reading the manuscript, K. Ronan for the help and patience in preparing the manuscript, and B. Kimmel for assistance with computer programs. This work was supported by the Deutsche Forschungsgemeinschaft and an American Cancer Society Fellowship, California Division (H.-M.J.), and by the National Cancer Institute (R.T.). The sequence of hUBF has been submitted to the EMBL/GenBank data libraries.

G_o is a major growth cone protein subject to regulation by GAP-43

Stephen M. Strittmatter*, Dario Valenzuela*, Timothy E. Kennedy†, Eva J. Neer‡ & Mark C. Fishman*§

* Developmental Biology Laboratory, Massachusetts General Hospital; Departments of Neurology and Medicine, Harvard Medical School; Howard Hughes Medical Institute, Boston, Massachusetts 02114, USA

† Center for Neurobiology and Behavior, Columbia University, 722 W. 168th Street, New York, New York 10032, USA

‡ Cardiovascular Division, Brigham and Woman's Hospital, 75 Francis Street, Boston, Massachusetts 02115, USA

G_o, a GTP-binding protein that transduces information from transmembrane receptors, has been found to be a major component of the neuronal growth cone membrane. GAP-43, an intracellular growth cone protein closely associated with neuronal growth, stimulates GTP-γ-S binding to G_o. It does so through an amino-terminal domain homologous to G-linked transmembrane receptors. Thus, G_o in the growth cone may be regulated by intracellular as well as extracellular signals.

THE complex state of neural connectivity achieved during brain development requires highly selective neurite extension. The

distal tip of neuronal processes, a structure termed the growth cone, is critical for transduction of extracellular signals into directed growth^{1,2}. It has been proposed that the levels of certain intraneuronal proteins also modulate growth. GAP-43, a neuron-specific protein, is the best characterized of these 'growth-associated proteins' (GAPs) whose expression is closely correlated with axonal growth^{3,4}. GAP-43 is tightly associated with the inner surface of the plasma membrane, especially at the growth cone^{5–9}. Although its normal action *in vivo* is unclear, GAP-43 can bind calmodulin, inhibit phosphatidylinositol phosphate kinase (PIP kinase), be phosphorylated by protein kinase C, affect neurotransmitter release and enhance filopodia in non-neuronal cells^{3,4,10–14}.

An important aspect of morphogenesis for any cell is the integration of intracellular and extracellular growth signals. As the neuronal growth cone membrane contains only a few predominant proteins¹⁵, it provides the opportunity to identify molecules involved in the coordination of various growth sig-

§ To whom correspondence should be addressed.

nals. Therefore, we sought to characterize the prominent unidentified proteins of the growth cone membrane and to examine interactions among its constituents. We find that G_o , a member of the signal-transducing G-protein family¹⁶⁻¹⁸, is one of the main components of the growth cone membrane. The binding of GTP- γ -S to G_o is regulated by GAP-43. Thus, G-protein transduction may be regulated by intracellular signals.

G_o is a major growth cone component

A growth cone membrane fraction was prepared from neonatal rat brain^{15,19}. This preparation has a simple protein composition as judged by SDS-PAGE (ref. 19 and Fig. 1). The most intensely stained band, migrating with a relative molecular mass (M_r) of 50,000–55,000, has been identified as tubulin¹⁹. Actin and GAP-43 co-migrate in the prominent band at M_r 43,000. The two other most prominent proteins have M_r values of about 35,000 and 40,000, and have been termed p34 and p38 (ref. 19). Their specific enrichment in the growth cone membrane is well documented¹⁹. These are the proteins we characterized further.

Proteins p34 and p38 have M_r s similar to those of the α - and β -subunits of the GTP-binding protein, G_o (refs 16–18, 20). Co-electrophoresis of the growth cone membranes with purified bovine brain G_o shows that p34 co-migrates with the β -subunit,

and p38 with the α -subunit of G_o (Fig. 1a). On immunoblotting, p34 reacts with an anti- β antiserum, and p38 reacts with an anti- α_o antiserum (α_o staining shown in Fig. 1b). Equal protein concentrations of p38 and α_o , as determined by Coomassie blue staining, exhibit identical immunoreactivity, showing that the predominant protein species of p38 must be α_o (Fig. 1b). The same approach shows that p34 is the β -subunit of G_o (data not shown). The α_i -subunit is about 10-fold less reactive than α_o with this antiserum²⁰, so it cannot account for much of the α_o immunoreactivity.

To verify this immunological and electrophoretic data, we obtained partial protein sequences from electrophoretically purified p34 and p38. Both p34 and p38 are blocked at their amino terminals. Sequences were obtained from tryptic fragments separated by HPLC and from fragments of a partial digest with *Staphylococcus aureus* V8 protease, separated by SDS-PAGE. The sequence for each of three peptides from p38 is identical to that of α_o , confirming that α_o is the chief component of p38 (Fig. 2). Other known α -subunits have similar but distinct sequences. The three p34 peptides have a sequence identical to that of the β -subunit of G proteins. Two peptides are from regions where β_1 - and β_2 -subunits are identical and the third contains a mixture of the sequences for β_1 and β_2 . Thus, the

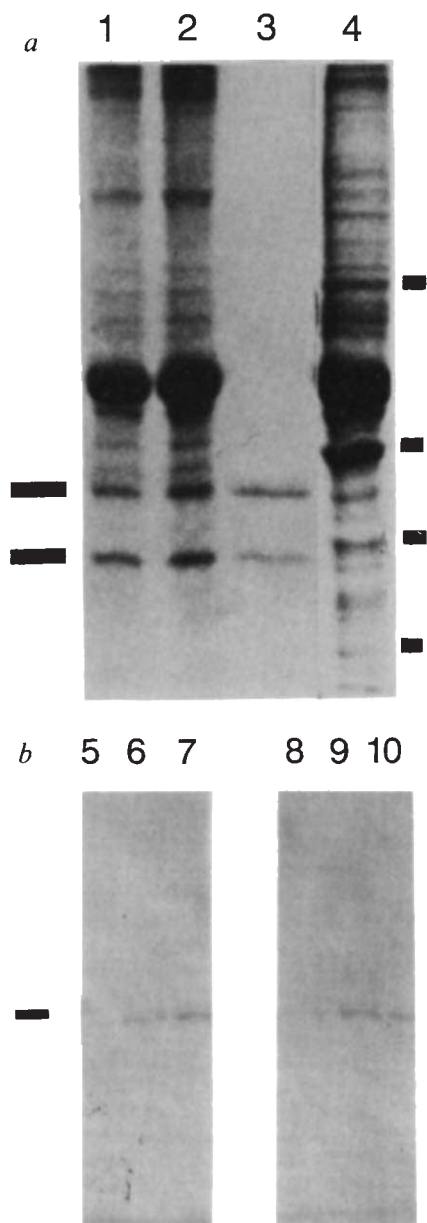


FIG. 1 Proteins of the axonal growth cone membrane. *a*, Growth cone membranes (two separate preparations, lanes 1 and 2 containing 8 and 12 μ g total protein respectively), purified bovine brain G_o (lane 3, containing 2 μ g protein), and crude brain homogenate (lane 4, containing 20 μ g protein) were electrophoresed through a 10% polyacrylamide gel in the presence of SDS and stained with Coomassie blue. Two bands, termed p34 and p38 are prominent (bars). These proteins co-migrate with the α - and β -subunits of purified G_o . In these conditions, actin and GAP-43 co-migrate with an apparent M_r of 43,000. The mobility of M_r markers of 66,000, 45,000, 36,000 and 29,000 (top to bottom) is shown by the squares. *b*, α_o immunoreactivity migrates at the position of α_o Coomassie blue staining (bar) in both the purified G_o preparation (lanes 5–7) and the growth cone membrane preparation (lanes 8–10). The gels were loaded so that Coomassie blue labelling of lane 5 of α_o was identical to that of p38 of lane 8, and similarly matched for the pairs 6 and 9, and 7 and 10. (Coomassie-stained gels (not shown) were run in parallel.) The pairs are immunostained to the same degree, as the total protein is increased from left to right, showing that p38 is as immunoreactive as authentic α_o with this antiserum. A small amount of immunoreactivity migrates just above α_o in the G_o sample which is probably due to slight contamination and cross-reaction with α_i . This is not seen in the growth cone membrane lanes. Similar experiments confirmed that p34 reacts with anti- β antiserum. The mobility of M_r markers of 116,000, 84,000, 58,000, 48,500, 36,500 and 26,600 (top to bottom) is shown by the squares. **METHODS.** Growth cone membranes were prepared with minor modifications from previous methods^{15,19}. Sprague-Dawley rats less than 24-h old were decapitated and the brains were homogenized at 4 °C with six passes in a glass-Teflon homogenizer in five volumes of 0.32 M sucrose, 1 mM Tris-HCl, 1 mM MgCl₂, pH 7.6. The following protease inhibitors were used throughout: 100 μ g ml⁻¹ soybean trypsin inhibitor, 1 μ g ml⁻¹ pepstatin A, 30 μ M leupeptin, and 1 mM phenylmethylsulphonyl fluoride. The crude brain homogenate was layered over a step gradient of sucrose at 0.75 M, 1.0 M and 2.2 M. The gradient was centrifuged at 250,000g for 40 min, and the 0.32/0.75 M interface was collected as the growth cone particle fraction. This fraction was lysed in 5 mM Tris-HCl, pH 7.6, and the membranes were collected by centrifugation at 250,000g for 40 min, washed by resuspension in 20 μ g ml⁻¹ saponin and 0.3 M Na₂SO₃ and again collected by centrifugation. Bovine brain G_o was prepared as described in ref. 33. Anti-bovine brain α_o and anti- β antisera were produced in rabbits²⁰ and characterized. Immunoblot samples were electrophoresed through 10% polyacrylamide gels with SDS and then electrophoretically transferred to nitrocellulose. Nonspecific protein binding sites were blocked with 10 mg ml⁻¹ BSA and the blots incubated with 1:400 anti- α_o antiserum or 1:100 anti- β antiserum²⁰ overnight at 4 °C. Bound antibody was detected by the avidin-biotin complex method (Vectastain) using tetrabenzidine as a peroxidase substrate.

α - and β -subunits of G_o are substantial constituents of the growth cone membrane subcellular fraction. We did not perform electrophoresis under conditions designed to detect the low M_r γ -subunit of G_o .

Although these subcellular fractions are substantially enriched in growth cone membrane, they are not pure¹⁵. We confirmed the presence and enrichment of G_o in growth cones by immunohistology. Previous immunohistology has shown that G_o is concentrated in the neuropil of adult rat brain²¹, that a related G protein, G_{olf} , is localized to the terminal region of primary olfactory neurons in the adult²², and that G_o stains throughout cultured primary neurons but is concentrated at regions of cell-cell contact²³. These studies are consistent with G_o localization to synaptic regions, but synapses and growth cones have different constituents^{1,2,15,19,57}. We therefore examined G_o distribution in PC-12 cells which had been treated with nerve growth factor to induce extension of processes tipped with easily seen growth cones. The most specific growth cone marker protein, GAP-43, is enriched in the growth cone of PC-12 cells as it is in those of primary neurons⁷. Immunofluorescence due to α_o is highly concentrated (but not found exclusively) in the growing tips of PC12 cells (Fig. 3).

GAP-43 stimulates GTP- γ -S binding to G_o

With the major growth cone membrane proteins identified, we wished to determine whether these components could form intermolecular complexes. In particular, G_o and GAP-43 were examined because they are the predominant noncytoskeletal proteins in the growth cone membrane. Attempts to isolate a GAP-43/ G_o complex physically by gel exclusion chromatography, sucrose density gradient centrifugation, immunoprecipitation and affinity chromatography were unsuccessful.

To identify transient GAP-43/ G_o interactions in solution, we measured [³⁵S]GTP- γ -S binding to purified G_o (refs 20, 24) in the presence of various concentrations of purified GAP-43. GAP-43 itself does not bind [³⁵S]GTP- γ -S, but GAP-43 stimulates specific [³⁵S]GTP- γ -S binding to G_o to a level $310 \pm 40\%$ ($n = 7$) of control (Fig. 4a). This effect is saturable and is half maximal in the presence of 150–800 nM GAP-43. If G_o and GAP-43 were free in solution and not membrane-bound *in vivo*, then their concentrations would be about 2 μ M in whole brain. Therefore, the affinity of GAP-43 for G_o measured in our assays is consistent with *in vivo* conditions. All assays were conducted in the presence of 200 μ g ml⁻¹ BSA, so a nonspecific protein effect by GAP-43 cannot explain the stimulation of GTP- γ -S binding. G_o is known to be partially inactivated during pre-incubation at 30 °C without GTP- γ -S (ref. 27). GAP-43 does not affect the degree of G_o thermal instability (data not shown). Although assessed in different conditions, ligand-receptor complexes which interact with G proteins stimulate GTP- γ -S binding to roughly the same extent as does GAP-43 (ref. 25).

GAP-43 has previously been shown to bind calmodulin, an interaction which is enhanced in the absence of calcium¹⁰. The physiological relevance of this association is unclear. The addition of Ca²⁺-calmodulin or EGTA-calmodulin to G_o causes no change in the ability of GAP-43 to stimulate [³⁵S]GTP- γ -S binding (data not shown).

GAP-43 decapeptide interacts with G_o

To determine which regions of GAP-43 are critical in stimulating GTP- γ -S binding to G_o , a series of synthetic peptides from the GAP-43 sequence were tested. A peptide containing the first 24 amino acids of GAP-43 stimulates GTP- γ -S binding in a saturable manner with an effector concentration for half-maximum response (EC_{50}) of 20 μ M (Fig. 4b). By contrast, peptides from three other regions of GAP-43 are ineffective even at higher concentration. The amino-terminal peptide stimulates [³⁵S]GTP- γ -S binding to the same level as does GAP-43 itself, and the addition of GAP-43 protein together with amino-terminal peptide does not stimulate binding to higher levels.

This suggests that the peptide is a full competitive agonist for GAP-43. The affinity of the peptide is approximately 100-fold lower than that of GAP-43. This may be due to the contribution of other segments of GAP-43 to G_o stimulation or to differences in secondary structure between the peptide and the intact protein.

Previously, we showed that the first 10 amino acids of GAP-43 are necessary and sufficient for its membrane association, and that the two cysteines at positions 3 and 4 are critical⁸. When a peptide containing just the first 10 amino acids of GAP-43 is added to G_o , [³⁵S]GTP- γ -S binding is inhibited by $80 \pm 6\%$ ($n = 4$) of control levels at saturating concentrations of peptide (Fig. 4c). The 50% inhibitory concentration (IC_{50}) is similar to the EC_{50} of the 1–24 peptide. The saturability of this effect suggests that the 1–10 peptide interacts with a specific site on G_o . To assess whether the peptide and GAP-43 compete for a single site, we compared the ability of GAP-43 to stimulate G_o in the presence and absence of the 1–10 peptide. The maximal level of [³⁵S]GTP- γ -S binding achieved in the presence of GAP-43 is not altered by the 1–10 peptide, but higher concentrations of GAP-43 are required in the presence of the peptide (Fig. 4d, EC_{50} of 12 μ M compared with 600 nM). Thus, the 1–10 peptide seems to be a reverse agonist for GAP-43, competing for the same site but causing an opposite effect. This localizes part of the binding site for G_o in GAP-43 to these 10 amino acids, but suggests that other regions are required for high affinity G_o /GAP-43 interaction and for a stimulatory effect on GTP- γ -S binding.

p38-A Alpha _o 11–23	IIHEDGFSGEDVK IIHEDGFSGEDVK
p38-B Alpha _o 70–85	MEDTEPFSAEILLSAMM MEDTEPFSAEILLSAMM
p38-C Alpha _o 119–132	IGAADYQPTTEQDIL IGAADYQPTTEQDIL
p34-A Beta ₁ 79–89 Beta ₂ 79–89	LIIWDSYTTNK LIIWDSYTTNK LIIWDSYTTNK
p34-B Beta ₁ 173–190 Beta ₂ 173–190	TGQQT ^{TT} _{VF} ^T _A GH ^T _S GDVMSL TGQQT ^{TT} _{VF} ^T _A GH ^T _S GDVMSL TGQQT ^{TT} _{VF} ^T _A GH ^T _S GDVMSL
p34-C Beta ₁ 338–340 Beta ₂ 338–340	IWN IWN IWN

FIG. 2 Partial protein sequence for p34 and p38 is identical to G_o . The sequence of three peptides from p38 matches the sequence of three peptides from α_o from rat brain⁴⁰. The sequence of three peptides from p34 is compared with that of the β_1 - and β_2 -subunits from bovine brain⁴¹. Note that two peptides are identical to regions in which β_1 and β_2 are identical. The other peptide contains a mixture of the sequences for β_1 and β_2 .

METHODS. Growth cone membranes were fractionated as in Fig. 1. When proteins were transferred to polyvinylidene difluoride (PVDF) membranes, the spots for p34 and p38 yielded no sequence, presumably because the proteins are amino-terminally blocked. Therefore, we obtained protein sequence from proteolytic fragments for p34 and p38. For tryptic digestion, the proteins were transferred to nitrocellulose, stained with Ponceau S and the appropriate bands were excised⁴². At the Harvard Microchemistry Facility, tryptic digestions were performed on the nitrocellulose membranes and the released peptides were separated by reverse phase HPLC and sequenced on a gas phase automated sequencer. Further amino-acid sequence was obtained for p34 following partial digestion with *S. aureus* V8 protease (Boehringer). Polyacrylamide cubes containing p34 were digested *in situ* with V8, fractionated by SDS-PAGE, electroblotted to PVDF membrane (Millipore) and visualized with Coomassie blue⁴³. Peptide fragments were excised and sequenced on an Applied Biosystems 470A gas phase sequencer at the Howard Hughes Medical Institute Protein Chemistry Core Facility of Columbia University to yield sequence p34-B.

To determine whether the cysteines of the amino-terminal decapeptide, which are critical for membrane association⁸, are also required for G_o regulation, we synthesized a decapeptide with threonines substituted for the two cysteines. This peptide has no effect on [³⁵S]GTP- γ -S binding to G_o , showing the specificity of the 1–10 peptide effect and the necessity of the cysteines (Fig. 4c).

The other group of proteins known to stimulate GTP- γ -S binding to G proteins are hormone and neurotransmitter receptors. Overall receptor structure is much different from that of GAP-43. We searched for homologies between these proteins, focusing our search on those domains thought to interact with G proteins. For the receptors, site-directed mutagenesis and peptide competition studies have implicated the carboxyl end of the third cytoplasmic loop, and the proximal end of the cytoplasmic tail in G-protein coupling^{26–28}. β_2 -adrenergic receptor linkage with G_s is interrupted most specifically by a point mutation at a palmitylated cysteine in the cytoplasmic tail of the protein²⁹. The cysteines in the amino terminus of GAP-43 which are required for G_o regulation are also subject to palmitylation *in vivo*⁹. A comparison of the GAP-43 amino terminus with the amino-terminal portion of the cytoplasmic tail from a series of receptors reveals a consensus sequence of hydrophobic-Leu-Cys-Cys-X-basic-basic (Fig. 5). Within this sequence, the cysteines, where studied, undergo palmitylation^{9,29,30}.

Other endogenous regulators of GTP- γ -S binding to G proteins are unknown. But the peptide toxin mastoparan reversibly activates G proteins³¹. GAP-43 is not homologous to this peptide.

Discussion

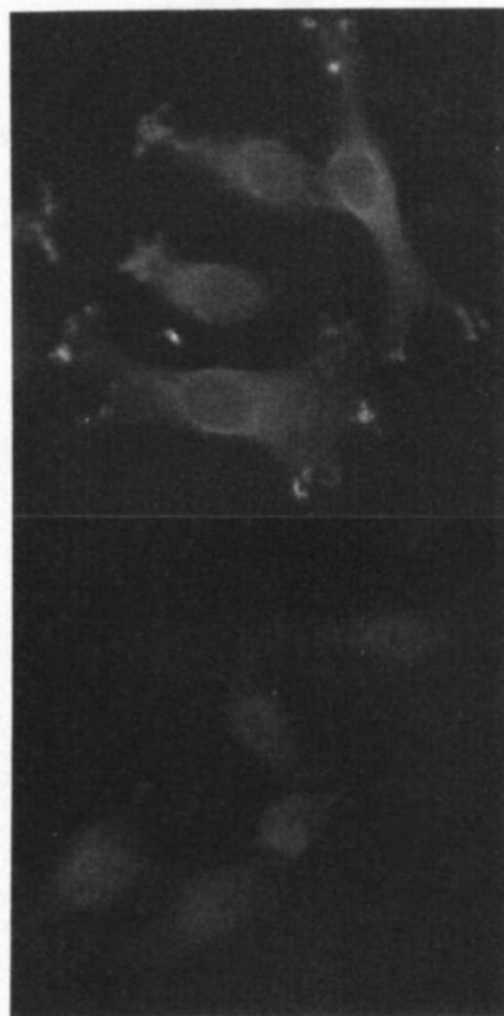
Our data show that G_o exists in strikingly high levels in the growth cone membrane, and can be regulated by an intracellular growth cone protein, GAP-43. These experiments do not directly assess the functional consequences of this G_o localization. In systems where the functions of G proteins are clearly defined, however, they are links between transmembrane receptors and ion channels or enzymes which modulate intracellular second messengers^{16–18}. Among G proteins, G_o is predominantly expressed in brain, where it is the most abundant G protein^{32,33}. Our data localize G_o to the tips of neurites in growing cells where it is the predominant noncytoskeletal protein. This prominence suggests a role for G_o in neurite regulation.

The receptors and effectors specifically linked to growth cone G_o remain to be elucidated. However, there is some evidence that G proteins transduce extracellular signals in the growth cone. In certain heliosome neurons, serotonin, which acts through a G-coupled receptor, is a potent inhibitor of neurite extension³⁴. Antibodies to the growth cone-localized proteins, N-CAM and L1, alter calcium levels and phosphatidylinositol metabolism in PC12 cells^{35,36}. The effect of these antibodies is blocked by pertussis toxin, a G_o/G_i inhibitor³⁶.

GAP-43 is unique among regulators of the heterotrimeric G proteins in that it is an intracellular protein. The potential for regulation of the heterotrimeric G proteins by intracellular macromolecules adds a new dimension to this membrane-transduction system. It remains to be seen whether the GAP-43 effect is specific for G_o , whether other G protein regulators exist and,

FIG. 3 α_o immunoreactivity is concentrated in the tips of PC12 processes. α_o staining of PC12 cells differentiated with nerve growth factor reveals high concentrations of the antigen at the distal tips of the cellular processes (top). There is also a lower level of diffuse staining in the region of the cell body surrounding the nucleus. The specificity of the antibody has been demonstrated (refs 20, 21, and Fig. 1b), but as a further control (bottom), identical samples were prepared with the addition of excess purified bovine brain G_o ($30 \mu\text{g ml}^{-1}$) to the incubation with antiserum, or with the substitution of normal rabbit serum for antiserum. These controls showed essentially no staining of the cellular processes.

METHODS. PC12 cells were grown for 48 h on coverslips treated with poly-D-lysine in the presence of 100 ng ml^{-1} NGF. Cells were fixed with 3.7% formaldehyde in PBS and then permeabilized with 0.1% Triton X-100. After incubation with 5 mg ml^{-1} BSA in PBS, cells were incubated with 1:1000 anti-bovine brain α_o antiserum for 1 h at 23°C , rinsed with PBS, and incubated with 0.3% H_2O_2 for 15 min to reduce background. Bound rabbit immunoglobulin was detected by use of the Texas-red conjugated donkey anti-rabbit IgG (Jackson Immunologicals).



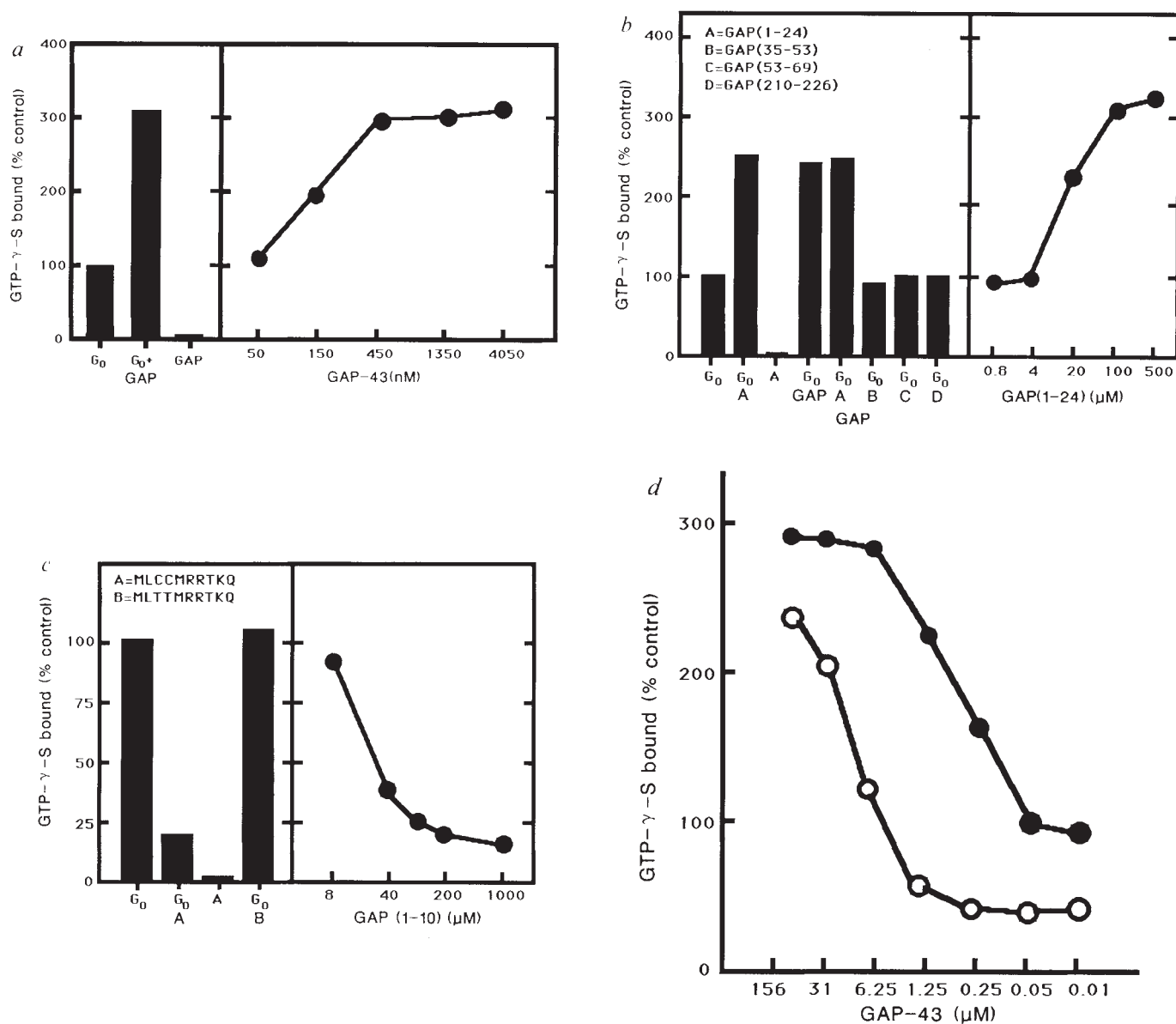


FIG. 4 Modulation of [35 S]GTP- γ -S binding to G_0 by GAP-43 and GAP-43 peptides. Specific [35 S]GTP- γ -S binding to G_0 is stimulated to $310 \pm 40\%$ ($n=7$ with this G_0 and this GAP-43 preparation) of control in the presence of $1 \mu\text{M}$ GAP-43 (a, left). The same concentration of GAP-43 without G_0 binds no detectable [35 S]GTP- γ -S. Variation of the GAP-43 concentration shows an EC_{50} of 150 nM for this effect (a, right). With different preparations of GAP-43 the range of EC_{50} was $150\text{--}800 \text{ nM}$ (compare d), perhaps reflecting partial inactivation of GAP-43 during purification. b (left), Addition of 1 mM peptide containing the first 24 amino acids from GAP-43, MLCCMRRTKQVEKNDEQKIEQDG, stimulates binding to the same level as with $1 \mu\text{M}$ GAP-43 (250%), and the addition of both GAP-43 and GAP (1-24) (A) causes no further stimulation. This peptide does not bind [35 S]GTP- γ -S in the absence of G_0 . Three other GAP-43 peptides, B, C and D (respectively GAP(35-53), ATKIQASFRGHITRKLLKD; GAP(53-69), DEKKGDAPEAEAEAKEK; and GAP(210-226), ARQDEKEDPEADQEHA) have no effect on [35 S]GTP- γ -S binding to G_0 at 1 mM . The stimulation of binding by the 1-24 peptide is saturable with an EC_{50} of $20 \mu\text{M}$ (right). c, Peptide consisting of the first 10 amino acids of GAP-43, MLCCMRRTKQ, inhibits [35 S]GTP- γ -S binding to G_0 by 80% at 1 mM (left). A peptide with threonine in place of cysteine at positions 3 and 4 has no effect on [35 S]GTP- γ -S binding to G_0 . The IC_{50} for the 1-10 peptide is $20 \mu\text{M}$ (right). d, [35 S]GTP- γ -S binding to G_0 as a function of GAP-43 concentration was determined in the presence (—○—) or absence (—●—) of $50 \mu\text{M}$ 1-10 peptide. GAP-43 reverses inhibition by the peptide, and the EC_{50} for GAP-43 is shifted from 600 nM to $12 \mu\text{M}$ by this concentration of peptide. This is consistent with direct competition between the 1-10 peptide and GAP-43.

METHODS. [35 S]GTP- γ -S (New England Nuclear, $2,000 \text{ Ci mmol}^{-1}$) binding

was determined as described previously^{20,24} by incubation of G_0 (1 nM , 10 ng) and [35 S]GTP- γ -S (2 nM) in $100 \mu\text{l}$ of 50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 6 mM MgCl_2 , 0.1% Lubrol PX, $200 \mu\text{g ml}^{-1}$ BSA, 1 mM DTT for 60 min at 30°C , filtration through nitrocellulose, and scintillation counting of bound radioactivity. The concentration of [35 S]GTP- γ -S in these assays is approximately one-tenth of the apparent K_D (refs. 20, 24), so that changes in both apparent K_D and B_{max} alter the level of binding. Total radioactivity bound was less than 10% of that added in all cases. Nonspecific binding was determined in the presence of $10 \mu\text{M}$ unlabelled GTP- γ -S and subtracted from total binding to determine specific binding. All assays were performed in duplicate. An average assay tube contained $800,000 \text{ c.p.m.}$ and yielded $30,000 \text{ c.p.m.}$ total binding and $1,500 \text{ c.p.m.}$ nonspecific binding. The addition of GAP-43 increased specific binding without effect on nonspecific binding. GAP-43 activated two different samples of G_0 and one of α_0 , all prepared in Lubrol PX. Using combinations of different GAP-43 and G_0 preparations, the level of GTP- γ -S binding varied between 100 and 425% of control, with an average of $230 \pm 15\%$ (s.e.m., $n=42$). Activation by GAP-43 seems sensitive to the detergent used in the purification of G_0 and the composition of the binding buffer. GAP-43 was purified by a modification of the method of Zwiers *et al.*⁴⁴, and was greater than 97% pure by Coomassie blue staining of SDS-PAGE gels. The protein was dialysed extensively against the assay buffer and added to the GTP- γ -S- G_0 incubations in the concentrations indicated above. Peptides were synthesized at the Howard Hughes Medical Institute laboratory, purified by reverse phase HPLC, and sequence was confirmed by amino-acid sequencing. Peptides were dissolved in assay buffer and added to the GTP- γ -S- G_0 incubation at the indicated concentrations.

GAP-43, Human	M L C C E M - R R
GAP-43, Fish	M L C C I - R R
B ₁ -adrenergic	L L C C A - R R
B ₂ -adrenergic	L L C L - R R
Rhodopsin	T I C C G - K N
Blue-opsin	M V I C G - K A
A ₁ -adrenergic	I L G C Q C R S
pA ₂ -adrenergic	I L C R G D K R
Mus-ACh-1	L L C R W - K R
Mus-ACh-2	L M C H Y - K N
Mus-ACh-3	L L C Q C D K K
Mus-ACh-4	L L C Q Y - R N
Mus-ACh-5	L L C R W - K K
5HT-1A	I I K C W F C R
5HT-1C	L R C D Y - K P
5HT-2	I Q C Q Y - K E

FIG. 5 Homology between GAP-43 amino terminus and the cytoplasmic tail of G-linked receptors. The first 10 amino acids of GAP-43 were compared with the sequence of the G-linked receptor family. Homology between the first seven amino acids of GAP-43 and the amino-terminal segment of the cytoplasmic tail of a number of G-linked receptors is shown. Residues which conform to the consensus sequence hydrophobic-Leu-Cys-Cys-X-basic-basic are shaded. The cysteines illustrated in the receptors are located approximately 13 amino acids distal to the last transmembrane region, correspond to Cys₃₄₁ of the β_2 -adrenergic receptor, and are aligned as in ref. 31. The sequences shown are from human⁴⁵ and fish⁴⁶ GAP-43, human β_1 -adrenergic receptor (B₁-adrenergic, ref. 47), human β_2 -adrenergic receptor (B₂-adrenergic, ref. 48), human rhodopsin (rhodopsin, ref. 49), human blue-opsin (blue-opsin, ref. 50), human α_1 -adrenergic receptor (A₁-adrenergic, ref. 51), human platelet α_2 -adrenergic receptor (pA₂-adrenergic, ref. 52), human muscarinic receptor subtypes 1-5 (Mus-ACh-1-5, refs 31, 53), human serotonin receptor subtype 1_A (5HT-1A, ref. 54), and rat serotonin receptor subtypes 1_C and 2 (5HT-1C, ref. 55, and 5HT-2, ref. 56).

if so, whether they resemble GAP-43. Several GTP-binding proteins related to the heterotrimeric G proteins by sequence and mechanism are subject to intracellular regulation. The

GTPase activity of normal Ras proteins is stimulated by a widely distributed 120K (M_r 120,000) intracellular protein, GAP (GTPase activating protein³⁷). (The similarity in nomenclature between GAP and GAP-43 is fortuitous.) Binding of GTP to the protein synthesis cofactors, eIF-2 and EF-Tu, is modulated by eIF-2B and EF-Ts, respectively^{38,39}.

The amino terminus of GAP-43 enhances GTP- γ -S binding to G_o. The first 24 amino acids of GAP-43 stimulate GTP- γ -S binding to the same level as GAP-43. A shorter peptide, 1-10, does not stimulate GTP- γ -S binding but does compete with GAP-43 for binding to G_o. Thus, the first 10 amino acids of GAP-43 are an important determinant of its ability to bind to G_o, but other portions of the amino terminus are required for the stimulation of GTP- γ -S binding to G_o.

The cysteines in the region of GAP-43 (ref. 9) and receptors^{29,30} which stimulate GTP- γ -S binding to G_o are subject to palmitoylation. In our experiments, the amino-terminal peptides and probably the GAP-43 (prepared by pH11 extraction of membranes) exist in a nonpalmitoylated state. The relative ability of palmitoylated compared with nonpalmitoylated GAP-43, and G-linked receptors, to stimulate G proteins is unknown. It has been proposed that rapid palmitoylation-depalmitoylation has a regulatory role^{9,29}.

Both GAP-43 and G-linked transmembrane receptors stimulate GTP- γ -S binding to G_o through discrete domains which share sequence homology. Although they might achieve this action through different mechanisms, the sequence homology argues for one binding site on G_o for both GAP-43 and transmembrane receptors. A likely interpretation is that GAP-43 mimics the cytosolic tail of transmembrane receptors in the activation of G_o, thereby creating a GTP-bound α -subunit which triggers an intracellular second messenger system. Alternatively, GAP-43 binding to G_o might function primarily to disrupt G_o-receptor or G_o-effector interactions. It is also conceivable that GAP-43 is an effector of G_o. In any of these models, growth cone G_o could coordinate extracellular signals with an intracellular programme for morphogenesis. □

Received 11 January; accepted 28 February 1990.

- Bray, D. & Hollenbeck, P. J. *A. Rev. Cell Biol.* **4**, 43-61 (1988).
- Locke, R. O. *Neuroscience* **20**, 719-729 (1987).
- Skene, J. H. P. *A. Rev. Neurosci.* **12**, 127-156 (1989).
- Benowitz, L. I. & Routtenberg, A. *Trends Neurosci.* **10**, 527-531 (1987).
- Meiri, K. A., Pfenninger, K. H. & Willard, M. B. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3537-3541 (1986).
- Skene, J. H. P. *et al. Science* **233**, 783-785 (1986).
- Van Hoff, C. O. M. *et al. J. Cell Biol.* **108**, 1115-1125 (1989).
- Zuber, M. X., Strittmatter, S. M. & Fishman, M. C. *Nature* **341**, 345-348 (1989).
- Skene, J. H. P. & Virag, I. *J. Cell Biol.* **108**, 613-624 (1989).
- Alexander, K. A., Cimlar, B. M., Meier, K. E. & Storm, D. R. *J. Biol. Chem.* **262**, 6108-6113 (1987).
- Van Dongen, C. J., Zwiers, H., De Graan, P. N. E. & Gispen, W. H. *Biochem. biophys. Res. Commun.* **128**, 1219-1227 (1985).
- Routtenberg, A. *Annu. N.Y. Acad. Sci.* **444**, 203-211 (1985).
- Dekker, L. V., De Graan, P. N. E., Oestreicher, A. B., Versteeg, D. H. G. & Gispen, W. H. *Nature* **342**, 74-76 (1989).
- Zuber, M. X., Goodman, D. W., Karns, L. R. & Fishman, M. C. *Science* **244**, 1193-1195 (1989).
- Pfenninger, K. H., Ellis, L., Johnson, M. P., Friedman, L. B. & Somio, S. *Cell* **35**, 573-584 (1983).
- Stryer, L. & Bourne, H. R. *A. Rev. Cell Biol.* **2**, 391-419 (1986).
- Gilman, A. G. *A. Rev. Biochem.* **56**, 615-649 (1987).
- Neer, E. J. & Clapham, D. E. *Nature* **333**, 129-134 (1988).
- Simkowitz, P., Ellis, L. & Pfenninger, K. H. *J. Neurosci.* **9**, 1004-1017 (1989).
- Huff, R. M., Axton, J. M. & Neer, E. J. *J. Biol. Chem.* **260**, 10864-10871 (1985).
- Worley, P. F., Baraban, J. M., Van Dop, C., Neer, E. J. & Snyder, S. H. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4561-4565 (1986).
- Jones, D. T. & Reed, R. R. *Science* **244**, 790-795 (1989).
- Brabet, P. *et al. J. Neurosci.* **8**, 701-708 (1988).
- Northrup, J. K., Smigel, M. D. & Gilman, A. G. *J. Biol. Chem.* **257**, 11416-11423 (1982).
- Florio, V. A. & Sternweis, P. C. *J. Biol. Chem.* **264**, 3909-3915 (1989).
- Konig, B. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 6878-6882 (1989).
- O'Dowd, B. F. *et al. J. Biol. Chem.* **263**, 15985-15992 (1988).
- Strader, C. D. *et al. J. Biol. Chem.* **262**, 16439-16443 (1987).
- O'Dowd, B. F., Hnatowich, M., Caron, M. G., Lefkowitz, R. J. & Bouvier, M. *J. Biol. Chem.* **264**, 7564-7569 (1989).
- Ovchinnikov, Y. A., Abdulaev, N. G. & Bogachuk, A. S. *FEBS Lett.* **230**, 1-5 (1988).

- Higashijima, T., Uzu, S., Nakajima, T. & Ross, E. M. *J. Biol. Chem.* **263**, 6491-6494 (1988).
- Sternweis, P. C. & Robishaw, J. D. *J. Biol. Chem.* **259**, 11517-11526 (1984).
- Neer, E. J., Lok, J. M. & Wolf, L. G. *J. Biol. Chem.* **259**, 14222-14229 (1984).
- Haydon, P. G., McCobb, D. P. & Kater, S. B. *Science* **226**, 561-564 (1984).
- Letourneau, P. C. & Shattuck, T. A. *Development* **105**, 505-519 (1989).
- Schuch, U., Lohse, M. J. & Schachner, M. *Neuron* **3**, 13-20 (1989).
- Trahey, M. & McCormick, F. *Science* **238**, 542-545 (1987).
- Hachmann, J., Miller, D. L. & Weissbach, H. *Arch. Biochem. Biophys.* **147**, 457-466 (1971).
- Safer, B. *Cell* **33**, 7-8 (1983).
- Itoh, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 3776-3780 (1986).
- Fong, H. K. W., Amatruda, T. T., Birren, B. W. & Simon, M. I. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3792-3796 (1987).
- Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E. & Kent, S. B. H. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6970-6974 (1987).
- Kennedy, T. E., Gawinowicz, M. A., Barzilai, A., Kandel, E. R. & Sweatt, J. D. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7008-7012 (1988).
- Zwiers, H., Verhaghen, J., van Dongen, C. J., de Graan, P. N. E. & Gispen, W. H. *J. Neurochem.* **44**, 1083-1090 (1985).
- Ng, S.-C., de la Monte, S. M., Conboy, G. L., Karns, L. R. & Fishman, M. C. *Neuron* **1**, 133-139 (1988).
- LaBate, M. E. & Skene, J. H. P. *Neuron* **3**, 299-310 (1989).
- Frielle, T., Collins, S., Daniel, K. W., Caron, M. G. & Lefkowitz, R. J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7920-7924 (1987).
- Koblika, B. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 46-50 (1987).
- Nathans, J. D. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4851-4855 (1984).
- Nathans, J., Thomas, D. & Hogness, D. S. *Science* **232**, 193-202 (1986).
- Cotecchia, S., Schwinn, D. A., Randall, R. J., Lefkowitz, R. J. & Caron, M. G. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7159-7163 (1988).
- Koblika, B. K. *et al. Science* **238**, 650-656 (1987).
- Peralta, E. G. *et al. EMBO J.* **6**, 3923-3929 (1987).
- Koblika, B. K. *et al. Nature* **329**, 75-79 (1987).
- Julius, D., MacDermott, A. B., Axel, R. & Jessell, T. M. *Science* **241**, 473-479 (1988).
- Pritchett, D. B. *et al. EMBO J.* **7**, 4135-4140 (1988).
- Bazier, L. & Fishman, M. C. *J. Neurosci.* **8**, 2305-2311 (1987).

ACKNOWLEDGEMENTS. We thank A. Gawinowicz for technical assistance and the Natural Sciences and Engineering Research Council of Canada for a postgraduate scholarship to T.E.K.

Sub-arcsecond observations of the solar X-ray corona

L. Golub*, M. Herant†, K. Kalata*‡, I. Lovas§,
G. Nystrom*, F. Pardo†, E. Spiller§ & J. Wilczynski§

* Smithsonian Astrophysical Observatory, 60 Garden St, Cambridge, Massachusetts 02138, USA

† Harvard University, Cambridge, Massachusetts 02138, USA

‡ Rosenstiel Center, Brandeis University, Waltham, Massachusetts 02154, USA

§ IBM, T. J. Watson Research Center, Yorktown Heights, New York 10598, USA

SOFT X-ray observations are recognized as the best way to study the solar corona, as they are largely free of contaminating emission from other temperature regimes. They provide the only available method for seeing the corona on the disk, thereby avoiding the line-of-sight integration effects which are troublesome in limb observations. Here we present results from a high-resolution multilayer-coated X-ray imaging telescope, part of the Normal Incidence X-ray Telescope sounding rocket payload. From a flight at 16:35 UT on 11 September 1989 we obtained 40 images of the solar X-ray corona, with spatial resolution up to 0.75 arcsec. Images of the peak of a two-ribbon flare showed detailed structure within each ribbon, as well as the expected bright arches of emission connecting the ribbons. Active regions showed structure at a scale of 0.75 arcsec that was not visible in our 2-arcsec images taken during the same flight. The number of X-ray bright points¹ is small, consistent with predictions based on the previous solar cycle². The topology of the magnetic structure is complex and highly tangled, implying that the magnetic complexity of the photosphere is paralleled in the corona.

The new telescope uses multilayer coatings to increase the X-ray reflectivity³⁻⁶. The telescope was configured as a 25-cm diameter, 200-cm focal length prime focus ($f/8$) instrument in the 1989 flight. The multilayer coatings on the telescope optics allow imaging of the Fe XVI and Mg X coronal emission lines near a wavelength of 63.5 Å with a passband of 1.4 Å (full width at half maximum). The images were made using modified Kodak T-max 400 and Technical Pan films.

Multilayer mirrors have been used on several flights of sounding rockets⁷⁻⁹ as well as on the Phobos interplanetary spacecraft¹⁰. These early flights demonstrated that the basic technique works without major problems and that image quality at least as good as that available with grazing incidence optics can be obtained with normal incidence multilayers. Our programme has focused largely on developing multilayer coatings that will image a spectral line or lines formed at temperatures in the range 1 to 4 million degrees, for studies of coronal active regions and their dynamics. The choices of wavelength are limited; we have selected 63.5 Å, which allows imaging of the Fe XVI and Mg X lines. The performance of the coatings used in the Normal Incidence X-ray Telescope (NIXT) rocket and the effective instrumental response to the solar spectrum have been discussed in detail elsewhere^{11,12}.

The optics for the first successful flight of this payload in 1988 were of Ritchey-Chretien design, consisting of a pair of hyperbolic mirrors in Cassegrain configuration¹³. This design permits sub-arcsecond image quality over a wide field of view and produces the image size necessary to match the resolution to the capabilities of photographic emulsions. But we were able at the time to produce only a low value of peak reflectivity (4%), which resulted in a low throughput in Ritchey-Chretien configuration, so that only weak, underexposed images were obtained during the five minutes of available observing time. For the most recent flight, we therefore changed to a prime focus system; this increased the effective collecting area by a factor of 25 by eliminating the second 4% reflection, and

increased brightness by another factor of 12 by going to a lower f -number ($f/8$ as opposed to $f/30$).

The data are recorded on 70-mm film, using a Hasselblad 500 EL/M camera that was modified and flight qualified for us by Hasselblad. For the 1989 flight we used two photographic emulsions to record the X-ray data. Special runs of T-max 400 and Technical Pan were manufactured for this programme by Kodak. The former is a high-speed film with high spatial resolution, while the latter has extremely high resolution and moderate sensitivity; both films were made without the usual protective gelatin top coat and with a clear antistatic backing. The resolution is not limited by the optics but by the film resolution, which was measured in our laboratory at soft X-ray wavelengths; the limits are ~ 2 arcsec for the fast film and 0.75 arcsec for the finer-grained emulsion. Because these films are sensitive to visible light as well as to X-rays, it is necessary to make filters that attenuate the visible by a factor of 10^9 while transmitting a reasonable fraction of the soft X-rays. It is also necessary to make large, unobstructed heat-rejection filters to cover the entrance aperture in order to prevent defocusing of the telescope during the flight¹⁴.

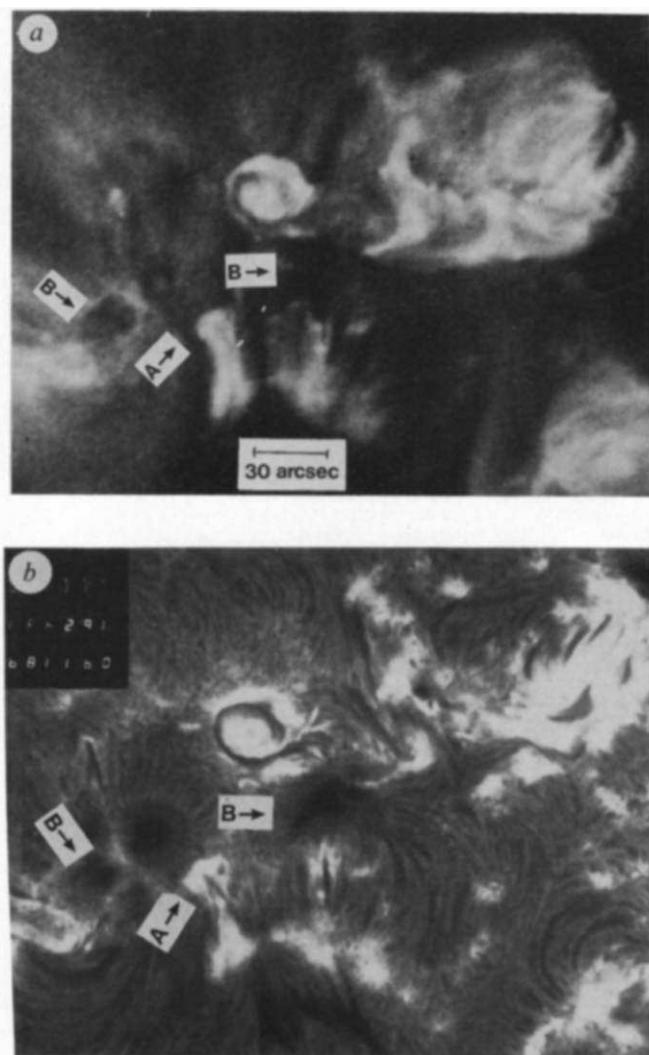


FIG. 1 *a*, Enlargement of a 60-s exposure NIXT soft X-ray image taken at 16:38 UT on 11 September 1989. Arrow labelled A points to footpoint of coronal loop; arrows labelled B indicate dark coronal regions located above sunspots. *b*, $H\alpha$ photo to same scale as *a*, taken at 16:24 UT. Arrow labelled A indicates penumbral brightenings associated with the footpoint of coronal loops; arrows labelled B indicate sunspots. (Photo courtesy Big Bear Solar Observatory).

The NIXT payload has flown three times to date. The first flight produced no measureable X-rays due to a filter transmission problem. In the second flight^{12,13} we recorded emission from a solar flare and marginal detection of active-region flux. The third flight, in prime focus optical configuration, resulted in detection of all known coronal X-ray features from active regions and flares to X-ray bright points and polar plumes; here we present a preliminary analysis of the data from this flight. In both of the successful flights we were fortunate in having enhanced X-ray brightness because of the presence of flare events. For the 23 June 1988 flight this occurrence was fortuitous; for the 11 September 1989 flight it was planned via monitoring of the soft X-ray intensity. However, we were again fortunate to observe the start of a second flare during our flight.

The NIXT payload was launched at 16:34:57 UT on 11 September 1989; the Sun was imaged (stable fine pointing) at 16:36:35 and loss of fine pointing occurred at 16:41:50 UT. During this time we obtained 38 exposures, ranging from 1 to 60 s; all of the exposures contain usable soft X-ray imaging. Our flight film was processed within hours of payload recovery, using the facilities at the National Solar Observatory (NSO/SPO): preliminary photographic processing was also carried out at NSO/SPO. Some images were digitized at NSO/SPO and were transferred to our group's data analysis facility at SAO. Data are stored in a dedicated microVax II supporting the ANA interactive data manipulation language (R. Shine, personal communication), with a Calcomp 1280X1024 image display system. In addition to the NIXT X-ray data we have also received longitudinal magnetograms, He I 10,830-Å full Sun scans (both from NSO/KPNO), high-resolution H α and video magnetograms (from Big Bear Solar Observatory) and patrol H α spectroheliograms (routine full-disk H α spectroheliograms taken by a number of stations in an automatic manner). These data sets have been coaligned using the limb which is clearly visible in the X-ray data.

The data show a marked increase in fine structure when the resolution is improved from 2 arcsec on our fast emulsion to 0.75 arcsec on the finer-grained film. There are loop structures visible at the coarser resolution level, as has been known since the early 1970s from grazing incidence imaging. However, the sudden appearance of significantly more structure at a resolution level of ~ 1 arcsec indicates that this preferred transverse scale size for loop structures may represent a fundamental parameter in theories for the formation of structure under conditions applicable to the solar corona. We note, however, that the images seem to suggest that structuring on still finer scales will be seen in future flights with improved spatial resolution.

Filaments appear in absorption in the NIXT data, which was not the case for the Skylab images in which filaments were optically thin. The $1/e$ absorption length for a typical filament at 63.5 Å is $\sim 1.5 \times 10^9$ cm; this is roughly 10 times smaller than the absorption length at 20 Å, which was the typical average wavelength in the broadband Skylab S-054 X-ray images. The filament cavities, which correspond to coronal regions of low X-ray emission surrounding filaments, are visible in our data just as they are in the shorter-wavelength grazing incidence images. There are also, however, locations—such as the southwest limb or some of the active regions near the centre of the disk—at which sharply outlined dark X-ray features coincide very precisely with H α filaments. Dark H α fibrils also appear in absorption against the bright X-ray features, which indicates that the cooler structure lies above the hotter coronal region, or that the fibrils somehow inhibit the heating mechanism.

Comparison of the active-region loops seen in X-rays with the regions visible in H α at the footpoints of the loops (Fig. 1) resolves a long-standing question by indicating that in general the loops terminate in the penumbrae of the spots, assuming of course that the active region has a spot with a penumbra (several examples are marked A in Fig. 1). This was the view advocated by Rosner *et al.*¹⁶, who argued that the transfer of turbulent

energy from the photosphere to the corona would be inhibited in spot umbrae, leading to reduced coronal heating. They also suggested that energy transfer in weak-field regions far away from sunspots would also fail to produce strong coronal heating because of the weakness of the field. A potentially significant feature of the observations is that there are penumbral brightenings visible in H α at precisely the places where clusters of coronal loops terminate A. The absence of coronal emission directly above sunspots B could also be explained by the presence of open, or effectively open, magnetic field lines. Even if the energy input through the footpoints of the magnetic field is then the same in the umbra and penumbra, the resulting appearance of heating in the corona would be substantially different; the sharpness of the umbra-penumbra boundary would then indicate the presence of a separatrix between open and closed field regions.

The brightest coronal feature observed in this flight is a small two-ribbon flare visible in H α northwest of the Sun's centre. This region brightened to $\sim C5$ (5×10^{-6} W m $^{-2}$ in the GDES 1–8 Å soft X-ray channel) in soft X-rays and we first observed it 2 minutes after the X-ray intensity peak. Figure 2 shows detail from a 10-s exposure image in which the two ribbons seen in H α are visible in X-rays and, as expected, bright arches are seen connecting the two ribbons. However, there are two major puzzles in the X-ray image: why are the remaining portions of the ribbons visible in X-rays without any loops connecting them at the coronal level, and, even more puzzling, why do the ribbons each have loop structure?

The X-ray image shows two bright arches across the neutral line at the southern end of the flare (A); this is expected from the standard model of such events (see, for example, ref. 17). It is clear, however, that the X-ray emission also extends northwards and that the two ribbons are visible separately without arches connecting them. Close examination shows that each ribbon contains loop structure (B, for example): the existence of loops within each flare ribbon has, as far we know, not been observed previously and is inconsistent with standard models for two-ribbon flares.

X-ray bright points (XBPs) are small, isolated bipolar magnetic features away from active regions¹. The number of X-ray bright points seen in this flight is very small: only two such regions are unambiguous, although three additional small faint features might also be considered. Of the two, the larger would probably be classified as a small active region. Both regions show definite loop structure, which is not easily seen in XBPs

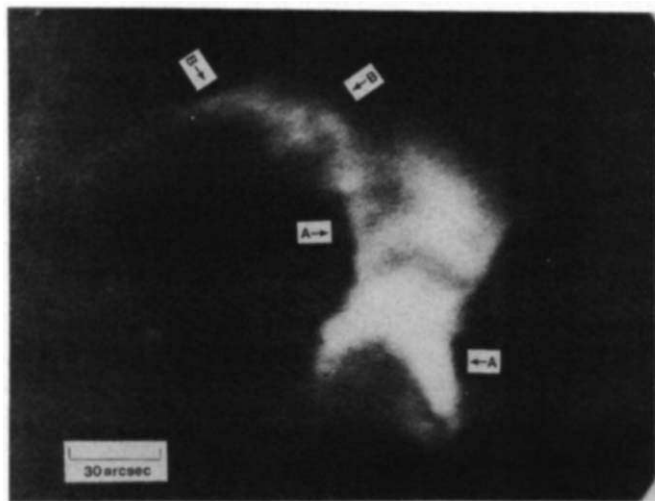


FIG. 2 Enlargement of a 10-s exposure NIXT image taken at 16:37 UT, showing details of the two-ribbon flare as seen in X-rays. Arrows labelled A point to the brightest features in the image, two arches spanning the neutral line; arrows labelled B indicate some of the loop structure seen in the individual flare ribbons.

with lower-resolution instruments. According to the empirical relation between sunspot number and number of XBPs which we derived during the previous solar cycle², there should have been twelve XBPs visible on this day; the expectation value for five XBPs assuming Poisson statistics is therefore low by 2σ . The instrumental sensitivity is such that the XBPs seen from Skylab would also have been seen on this flight had they been present. Taking into account the 90–100 XBPs that are typically seen at solar minimum², it is clear that this flight represents a solar maximum level of XBP activity.

We plan to fly the NIXT payload again in 1990, with no changes other than several minor modifications to increase the sensitivity. We are now able to make smoother multilayer coatings, which result in improvements by a factor of two to three in reflectivity at 63.5 Å (ref. 18), and we are developing new light-blocking filters with higher X-ray throughput. The combination of these two improvements will permit us to use a finer-grained (but slower) photographic emulsion and to observe quiet coronal structures with higher effective sensitivity than was available in the 1989 flight. □

Received 1 December 1989; accepted 27 February 1990.

1. Golub, L., Krieger, A. S., Silk, J. K., Timothy, A. F. & Vaiana, G. S. *Astrophys. J.* **189**, L93–L97 (1974).
2. Golub, L. *Phil. Trans. R. Soc. Lond. A* **297**, 595–604 (1980).
3. Spiller, E. *Appl. Phys. Lett.* **20**, 365–367 (1972).
4. Spiller, E. *Space Optics* 581–597 (National Academy of Sciences, Washington, DC, 1974).
5. Barbee, T. W. Jr. *AIP Proc.* Vol. 75 (eds Atwood, D. T. & Henke, B. L.) 131–145 (1981).
6. Spiller, E. *AIP Proc.* Vol. 75 (eds Atwood and Henke) 124–130 (1981).
7. Underwood, J. H., Bruner, M. E., Haisch, B. M., Brown, W. A. & Acton, L. W. *Science* **238**, 61–64 (1987).
8. Walker, A. B. C., Barbee, T. W. Jr., Hoover, R. B. & Lindblom, J. F. *Nature* **241**, 1781–1787 (1988).
9. Haisch, B. M., Whittemore, T. E., Joki, E. G. & Brookover, W. J. *Proc. SPIE* **982**, 38–45 (1988).
10. Sobelman, et al. Leningrad Institute Preprint No 241 (1988).
11. Spiller, E. & Golub, L. *Appl. Opt.* **28**, 2969–2974 (1989).
12. Golub, L. & Herant, M. *Proc. SPIE* **1160**, 629–635 (1989).

13. Golub, L., Nystrom, G., Spiller, E. & Wilczynski, J. *Proc. SPIE* **563**, 266–274 (1985).
14. Spiller, E., Grebe, K. & Golub, L. *Proc. SPIE* **1160**, 66–76 (1989).
15. Herant, M., Golub, L. & Neidig, D. *Solar Phys.* (in the press).
16. Rosner, R., Golub, L., Coppi, B. & Vaiana, G. S. *Astrophys. J.* **222**, 317–332 (1978).
17. Sturrock, P. A. (ed.) *Solar Flares* (Colorado Ass. Univ. Press, Boulder, 1980).
18. Spiller, E. *Proc. SPIE* **1160**, 271–279 (1989).

ACKNOWLEDGEMENTS. We thank K. Grebe for help with the soft X-ray prefilters, B. Kent for production of the light-blocking X-ray filter, and L. Kauder for use of a solar simulator for filter tests. L. Gilliam provided invaluable help with selection and testing of the flight films and with the film processing; we also thank T. Brown for darkroom help and L. November for microdensitometry. The NSO/Kitt Peak data used here are produced cooperatively by NSF/NOAO, NASA/GSFC and NOAA/SEL. This work was supported in part by a grant from NASA to the Smithsonian Institution.

Resonant internal waves in fluid flow

J. D. Pietrzak^{*†}, C. Kranenburg[†] & G. Abraham^{*}

^{*} Delft Hydraulics, PO Box 177, 2600 MH Delft, The Netherlands

[†] Department of Civil Engineering, Delft University of Technology, PO Box 5048, 2600 GA Delft, The Netherlands

UNDERSTANDING the flow of a stratified fluid over an uneven bed is an important problem in hydrodynamics. Previous field experiments using acoustic imagery^{1–3} have enhanced our understanding of internal-wave generation mechanisms. These studies have generally concentrated on isolated topographic features that show large-amplitude long internal waves and hydraulic jumps as the flow becomes critical defined in the hydraulic sense such that a long-wave speed vanishes with respect to a fixed coordinate system. We present field observations of stratified fluid flow over a nearly periodic topographic relief which show a unique series of resonant internal waves. The observations, carried out in the Rotterdam Waterway, show the presence of internal waves with wavelengths commensurate with those of the topography of the waterway bed and amplitudes that are three to four times larger than the relief of the topography generating them. Resonant internal waves may be important in understanding the flow dynamics of stratified coastal environments and may have consequences for atmospheric flows over mountain ranges.

Theoretical and experimental approaches^{4–9} to the general problem of stratified fluid flow over topography¹⁰ have provided useful models for atmospheric phenomena such as lee-wave generation on the downstream side of mountain ranges¹¹ and for oceanic phenomena such as tidal flow over sills^{1,12}. These studies have concentrated, however, on isolated topographic features rather than on periodic bottom topography. Here we consider the influence of an approximately periodic sequence of bottom ridges on the stratified flow. In contrast to lee waves, whose wavelength is governed by the background flow conditions, the wavelength of the trapped internal waves considered here is governed by the wavelength of the bottom ridges.

We present the results of a field survey carried out in the Rotterdam Waterway on 21 October 1987. The Rotterdam Waterway is a typical partially mixed estuary¹³ which discharges fresh water from the Rhine into the North Sea. Following an extensive echo-sounding and side-scan sonar survey, we selected a sampling location ~14 km from the mouth of the estuary, which has a series of thirteen pronounced bed ridges with typical wavelengths of 30–80 m and amplitudes of 0.5–1.0 m. Acoustic echo-sounding transects, using a 210-kHz transducer, were recorded over a 1-km section of the estuary. Velocity, conductivity and temperature profiles were recorded simultaneously from a moored vessel, and a standard algorithm was used to calculate density profiles¹⁴.

A sequence of five acoustic images, recorded over a 40-min period, showed the presence of large-amplitude internal waves on the decelerating tide as high-water slack was approached; the results are as follows:

Track number	Wavelength (m)	Amplitude (m)
1	80	1
2	50	2
3	40	4
4	30	4
5	30	3

The wave height of the largest internal waves was equal to nearly half the water depth. As a typical example, track 3 of the sequence is shown in Fig. 1. The largest waves, between $x = 250$ m and 400 m (where x is the horizontal distance scale), have amplitudes of 3.0–4.0 m and appear to be in phase with the topography, which has a dominant wavelength of ~40 m. The amplitude of the internal waves is three to four times larger than the amplitude of the topography. Large-amplitude waves are also observed elsewhere on the image, for example at $x = 680$ m, but the largest amplitudes correspond to ridges with a wavelength of ~40 m.

Tidal variations of the flow will influence the internal-wave response. If, however, the time variations occur on a scale much less than the period of the tide, the individual acoustic observations may be interpreted within the framework of quasi-steady flow¹⁵. In the following we use the quasi-steady approximation and linear internal-wave theory in order to find a physical interpretation of the resonant-like phenomenon seen in Fig. 1. We assume a steady, non-rotating, inviscid fluid. Small perturba-

[†] Present address: Department of Oceanography, University of British Columbia, Vancouver, British Columbia, Canada V6T 1W5

tions to a stable stratified shear flow, $U(z)$, can be analysed in terms of a single Fourier component of the vertical velocity, $w(x, z) = \hat{w}(z) e^{(ikx)}$. Under the Boussinesq approximation, $\hat{w}(z)$ satisfies an equation of the form¹⁶

$$\frac{d^2 \hat{w}}{dz^2} + \{I^2(z) - k^2\} \hat{w} = 0 \quad (1)$$

where

$$I^2(z) = \frac{N^2(z)}{U^2} - \frac{1}{U} \frac{d^2 U}{dz^2}$$

with $N(z)$ the buoyancy frequency. The surface was assumed to be a rigid lid, and a linearized boundary condition was used at the bed in which the topographic forcing was represented by a series of small-amplitude (h_0) sinusoidal features of wavenumber k . In this case the bottom boundary condition becomes $w = kU_0 h_0 \cos(kx)$, where U_0 is the horizontal velocity at the bed. Together with the boundary conditions, equation (1) forms a boundary-value problem for w . Equation (1) was integrated numerically using the observed velocity and density profiles. If, however, N and U are independent of depth, H , equation (1) can be integrated analytically to give

$$w(x, z) = \frac{U_0 k h_0 \sin[(I^2 - k^2)^{1/2}(H - z)] \cos(kx)}{\sin[(I^2 - k^2)^{1/2}H]} \quad (2)$$

Solutions of equation (1) have the same wavelength as the bottom configuration, and the wave amplitude varies with height. The important feature of the solution is that it is singular for values of the wavenumber $k = [I^2 - n^2 \pi^2 / H^2]^{1/2}$, where $n = 0, 1, 2, \dots$; that is, the conditions for resonance are satisfied if

the free internal wavelength that satisfies the unforced equations coincides with the wavelength of the topographic forcing.

The density and velocity profiles recorded during track 3 (Fig. 2) were used as input to a numerical scheme, which employed a fourth-order Runge-Kutta finite-difference integration technique. The acoustic reflections show vertical displacements, η , of lines of constant density from their undisturbed positions. In the linear approximation, $w = U d\eta/dx$ and solutions to equation (1) can be given in terms of η . In this way a direct comparison can be made with Fig. 1. The results of the numerical integration are shown in Fig. 3 for three cases, with topographic wavelengths of 35, 40 and 45 m and an amplitude in all cases of 1 m. The results suggest that the background flow conditions were such that the resonance condition was satisfied for a spectral component of the topography with a wavelength of 40 m.

A wavelength of ~ 40 m is observed in the topography between $x = 250$ and 400 m. Combined with the large-amplitude internal-wave response, the results suggest that the flow is near resonance for the observed internal waves. The actual topography seen in Fig. 1 could be represented as a superposition of individual Fourier components, so that the flow may be dominated by spectral components that most closely satisfy the resonance condition discussed above.

It is clear from the analytical solution (equation (2)) that the internal-wave response will depend on the local vertical structure of density and velocity. Therefore it is worth considering the influence of the velocity and density profiles on the solution of equation (1) as the flow decelerated from maximum flood. If the topography consisted of a monochromatic wavefield then, as the flow conditions changed, the resonance condition would

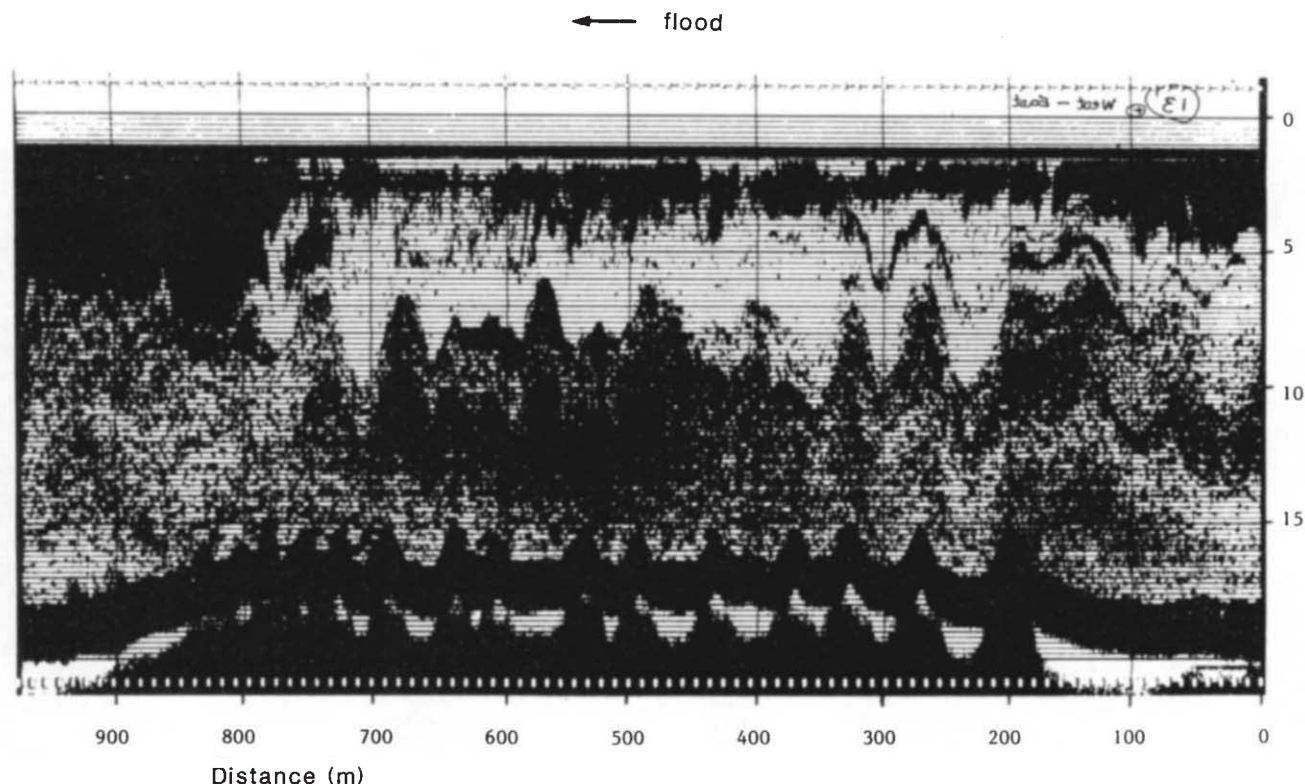
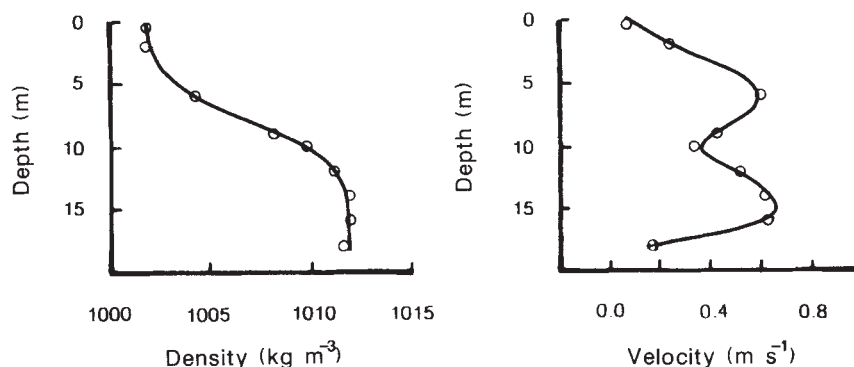


FIG. 1. 210-kHz acoustic image of track 3, showing a series of internal waves trapped over an approximately periodic sequence of bottom ridges. The horizontal scale is distance and the vertical scale is depth. The image was taken along the axis of the Rotterdam Waterway. The direction of the flood is from right to left. The acoustic reflections could plausibly have been produced by neutrally buoyant particles or salinity micro-structure. This was

verified by towing a conductivity probe through the water at a depth of 9 m below the surface and comparing the inferred vertical displacement of lines of constant density with the changes in density along the 9-m depth contour. Vertical lines running down the image were recorded at 100-m intervals using the position-fixing system on board the vessel. The largest-amplitude waves are observed between $x = 250$ and 400 m.

FIG. 2 Observed density (a) and velocity (b) profiles during the acoustic transect of Fig. 1. The density profile, which is representative of other density profiles recorded during the survey, has a tanh shape, corresponding to nearly homogeneous surface and bottom layers and an approximately linear region across the pycnocline. The density difference between the upper and lower layers increases slowly as the intrusion of saline water progresses, reaching a maximum as the flow changes from flood to ebb. Typically the velocity profiles recorded on the decelerating flood exhibited low near-surface velocities, whereas those recorded on the accelerating phase of the ebb did not. Reduced velocities were also observed at mid-depth on the flood phase only. To give a sufficient number of steps in the vertical scale for numerical stability, the density and velocity profiles were approximated by analytic functions. In the case of density a tanh function using a least-squares fitting program was used. The velocity profiles were approximated by a spline function that allowed for errors in the form of a constant standard deviation of 0.05 m s^{-1}



applied to each data point. A flow reduction was recorded at mid-depth, possibly indicating contamination of the mean velocity profile by the internal waves. However, sensitivity tests suggested that the numerical procedure was fairly insensitive to the fitting procedure. The analytic functions (continuous lines) show reasonable agreement with the data (circles).

no longer be met and the large-amplitude response would diminish. But the topography has contributions from different wavenumbers, which emphasizes the response at those wavenumbers for which resonance occurs. Because the free internal waves obey a dispersion relationship, one may expect the shorter-wavelength components to satisfy the resonance condition as the flood slowly decelerates.

Time dependence is evident in the sequence summarized in the table above. Initially the components of long topographic wavelength showed a large-amplitude response, but shorter-wavelength components became apparent as the flow decelerated. During this period the wave pattern was clearly related to the topography. As the tidal flow slowly decelerated, the internal

Froude number decreased from about 0.9 to 0.6 and the shorter-wavelength components were being forced selectively. Here the internal Froude number is defined as the ratio of the depth-averaged tidal velocity to the long-wave phase speed of an internal-wave mode. For example, Fig. 1 shows large-amplitude internal waves that dominate the flow between $x = 250$ and 400 m ; as the tide decelerated and shorter-wavelength components satisfied the resonance condition, however, the large-amplitude response diminished in this region. In contrast, large-amplitude ($\sim 3\text{--}4 \text{ m}$) waves in phase with the topography were observed on the following image (Fig. 4) between $x = 700$ and 800 m . Fig. 1 shows both a longer wavelength and a much smaller response at this location. Numerical analysis using the back-

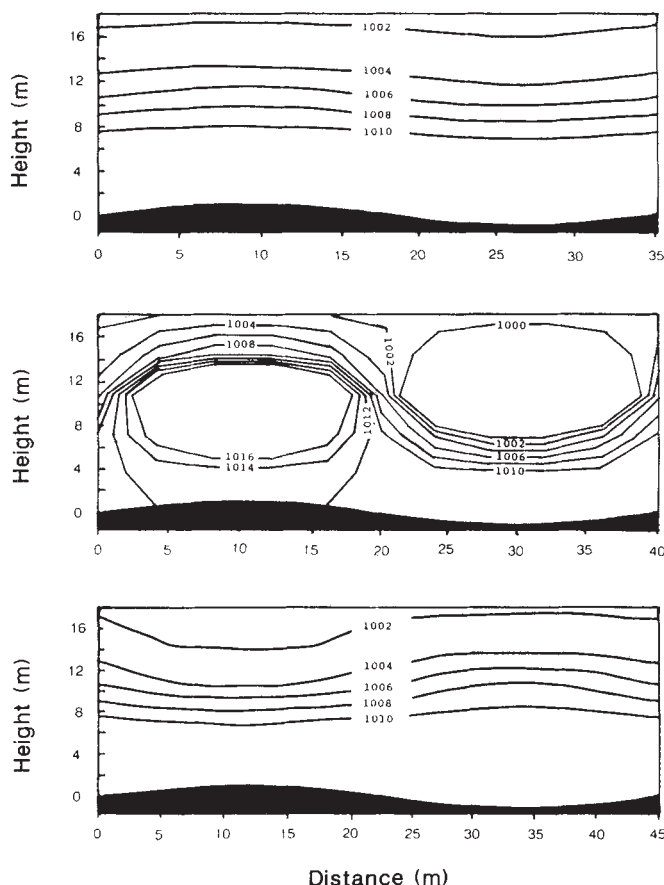


FIG. 3 Numerical solutions showing vertical displacements of lines of constant density. The bottom-topography wavelengths are (top to bottom) 35, 40 and 45 m. In all cases the amplitude of the topography is 1 m. The solution shows the response over one topographic wavelength. Although the numerical solution cannot deal with the singularity as resonance is approached (the solution begins to diverge from the sinusoidal form of the linear solution, developing regions of unstable density gradients), it does indicate that the resonance condition is being met for a narrow spectral bandwidth around a wavelength of 40 m. Within the experimental limits of the data set¹⁴ the numerical solution is reasonably consistent with the observation of large-amplitude waves associated with topographic wavelengths of about 40 m (Fig. 1). Although linear theory predicts the resonance condition, an alternative method is required to accurately predict wave amplitudes; in this respect a finite-amplitude theory is currently under investigation and results will be presented elsewhere.

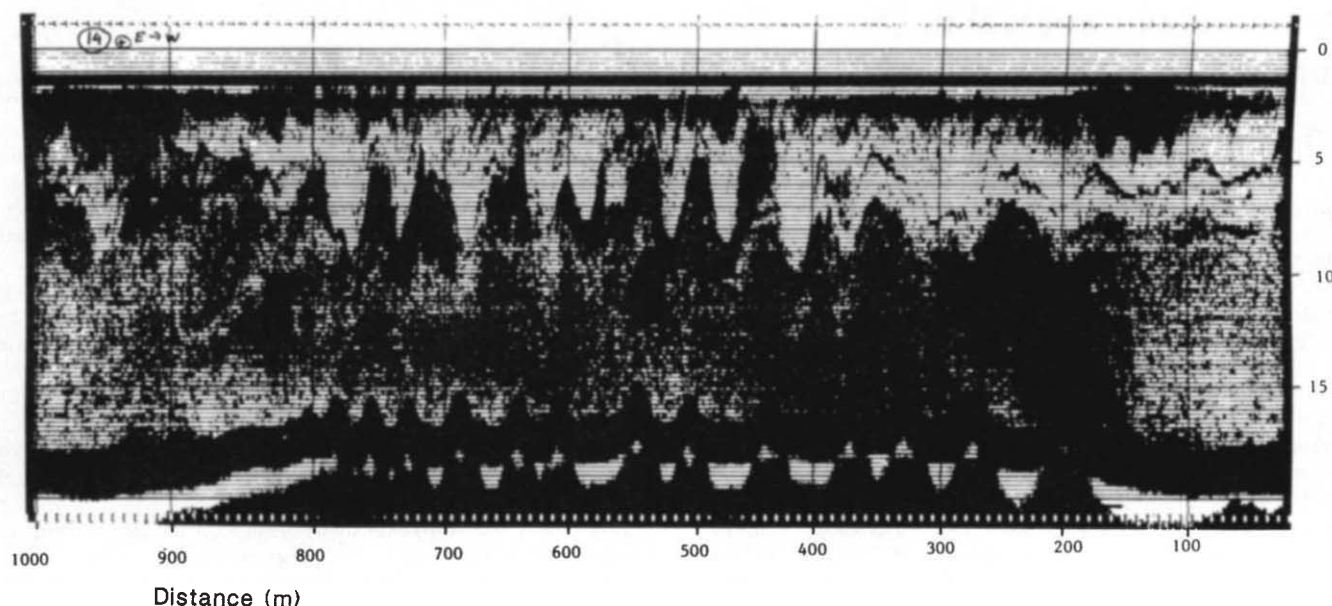


FIG. 4 210-kHz acoustic image showing track 4. The largest-amplitude waves are observed between $x=700$ and 800 m. The large-amplitude response

seen in Fig. 1 (track 3) at $x=250$ – 400 m has diminished.

ground flow data suggests that shorter-wavelength (~ 30 m) components satisfy the resonance condition here.

It is worth noting that internal-wave activity was recorded throughout the survey period when the flow was sub-critical, although the images recorded on the accelerating ebb showed considerable spatial and temporal variability. A flow reversal occurred at slack water and thereafter the trapped-wave system was replaced by waves of smaller amplitude that were not coupled uniquely to the topography. The difference between the response on the flood and on the ebb may be due to the reduced near-bed velocities on the ebb. The complete sequence, including a detailed assessment of the time dependence, is currently under investigation.

The internal waves that dominate the acoustic images show a clear relationship with the topography and suggest that the resonance condition can be isolated using a simple steady-state linear model. There are, however several limitations of this approach. The model can isolate only the wavelength that satisfies the resonance condition, and cannot be used to predict the amplitudes of the internal waves—this would require non-linear effects to be incorporated into the model. Another limitation is the applicability of the quasi-steady assumption for a tidal flow. Previous studies have shown that internal waves generated by stratified flow over bed features may be released during a tidal cycle and can travel either up or down the estuary, possibly interacting with features being generated elsewhere¹⁷. However, the largest amplitudes recorded on the acoustic image seem to occur when the resonance condition is satisfied, which suggests that the quasi-steady assumption is reasonable. Unsteadiness caused by either upstream influence¹⁸ or the time dependence of the tidal flow⁹ does not appear to interfere significantly with the resonance mechanism proposed here.

Previous field surveys in estuaries have revealed the presence of trapped internal waves¹⁷. We believe, however, that these are the first observations to show resonant internal waves over a wavy bed, the interpretation of which is supported by simple linear internal-wave theory. Because of the strong density gradients and the presence of bottom topography, estuaries may constitute one of the most favoured environments for producing resonance in the manner described. Mountain ranges are another possible source of this type of internal-wave generation, although in that case density stratification would have to provide

an effective upper boundary to reflect upward-propagating waves^{19,20}.

Internal waves trapped by the topography can have a significant effect on local flow dynamics. Recent studies^{21,22} suggest that internal waves of the observed amplitude increase the turbulent transport of momentum and mass and may need to be considered a sub-grid factor in numerical models. Further study of our data (to be presented elsewhere) also support this conclusion. It has been suggested that the greater bottom velocities induced by internal waves influence sediment transport and that standing internal waves may influence the formation of sand waves²³. Such interactions should be investigated, particularly when the flow is near resonance. In addition, internal waves can have significant implications for marine and estuarine biology, as local mixing could be increased by wave breaking. □

Received 13 November 1989; accepted 16 March 1990.

- Farmer, D. M. & Smith, J. D. *Deep Sea Res.* **17**, 239–254 (1980).
- Farmer, D. M. & Freeland, H. J. *Prog. Oceanogr.* **12**, 147–220 (1983).
- Chereskin, T. K. *J. geophys. Res.* **88**, 2649–2661 (1983).
- Long, R. R. *Tellus* **5**, 42–58 (1953).
- Long, R. R. *Tellus* **6**, 97–115 (1954).
- Long, R. R. *Tellus* **7**, 342–357 (1955).
- Baines, P. G. *J. Fluid Mech.* **82**, 147–159 (1977).
- Baines, P. G. *Tellus* **31**, 351–371 (1979).
- Maxworthy, T. *J. geophys. Res.* **84**, 338–346 (1979).
- Lamb, H. *Hydrodynamics* (Dover, New York, 1932).
- Smith, R. B. *J. Atmos. Sci.* **33**, 507–519 (1976).
- Farmer, D. M. & Smith, J. D. in *Hydrodynamics of Estuaries and Fjords* (ed. Nihoul, J. C. J.) (Elsevier, Amsterdam, 1978).
- Abraham, G. de Jong, P. & van Kruiningen, F. E. in *Lecture Notes on Coastal and Estuarine Studies* **16**, (ed. van de Kreeke, J.) 6–21 (Springer, Berlin, 1986).
- Pietrzak, J. D., Abraham, G. & Kranenburg, C. *Stratified Flows Rep. 6*, (Rijkswaterstaat, Delft Hydraulics, Delft University of Technology, 1989).
- Bell, T. H. *J. Fluid Mech.* **67**, 705–722 (1975).
- Gill, A. E. *Atmosphere-Ocean Dynamics* (Int. Geophys. Ser. 30) (Academic Press, New York, 1982).
- New, A. & Dyer, K. R. in *Physical processes in Estuaries* (eds Dronkers, J. & van Leussen, W.) 239–254 (Springer, Berlin, 1988).
- Grimshaw, R. H. J. & Smythe, N. *J. Fluid Mech.* **169**, 429–464 (1986).
- Bretherton, F. P. *Q. J. Met. Soc.* **95**, 213–243 (1969).
- Smith, R. B. *Adv. Geophys.* **21**, 87–230 (1979).
- Geyer, W. R. & Smith, J. D. *J. phys. Oceanogr.* **17**, 1668–1679 (1987).
- Kranenburg, C. *Estuar. Coastal Shelf Sci.* **27**, 15–32 (1988).
- Dyer, K. R. *Sedimentology* **29**, 885–889 (1982).

ACKNOWLEDGEMENTS. The field survey was carried out in co-operation with Rijkswaterstaat, from whom financial support was also obtained. We thank A. van der Wekken and B. Kranenburg for their co-operation. J. D. P. gratefully acknowledges support from a Royal Society fellowship. We thank T. Hibiya for critical reading of the manuscript.

Deep-water renewal in the northern North Atlantic

R. R. Dickson*, E. M. Gmitrowicz* & A. J. Watson†

* MAFF Fisheries Laboratory, Lowestoft, Suffolk NR33 0HT, UK

† Plymouth Marine Laboratory, West Hoe, Plymouth, Devon PL1 3DH, UK

THE renewal of the deep water of the world's oceans is accomplished through the input of dense water in both the northern North Atlantic and around Antarctica but direct measurement of the renewal rate has proved elusive. Here we describe the first successful long-term measurement of the transport of the northern inflow component where it passes south along the Continental Slope off East Greenland. There we observe a cold, bottom-intensified, long-slope flow with maximum speeds of around 100 cm s^{-1} , mean speeds of up to $25\text{--}30 \text{ cm s}^{-1}$, a dominant fluctuation timescale of a few days period but with little seasonal or interannual variation in speed or direction. We estimate the transport at densities (σ_θ) ≥ 27.80 to be stable at $10.7 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ and propose a circulation scheme for waters of this density throughout the northern gyre.

The global thermohaline circulation can be represented as a single overturning cell or 'vertical circulation' in which dense water produced by cooling at high latitudes sinks and spreads worldwide in the abyssal circulation, gradually upwelling as it does so until it returns to high latitudes in a system of surface return-flows. As deep-water formation releases a vast amount of heat to the atmosphere, it is thought to be significant that a shut-down of the overturning cell appears to have accompanied each abrupt switch of the ocean-atmosphere system towards glaciation¹. Today the abyssal circulation has the more practical importance of determining the location and timing of the peak dose to man arising from ocean-dumping of low-level radioactive (or other) wastes². For these and other reasons, the problem of quantifying the renewal of the deep waters of the world ocean remains "one of the major unsolved problems in oceanography"³.

The ventilation of the deep ocean worldwide is accomplished by the input of dense water from two high-latitude sources—those waters overflowing the Greenland–Scotland Ridge in the northern North Atlantic, and the outflow of dense water formed at various sites on the Continental Shelf of Antarctica during winter⁴. As these two inputs are together responsible for driving the global thermohaline circulation, numerous attempts have been made to assess their production rates. Direct measurement at the points of inflow has proved difficult, however, because these currents have proved too vigorous to maintain current-meter arrays for an adequate length of time without damage⁵.

Very recently, using a specially stiffened, low-drag mooring design based on 10-mm Kevlar cable, the MAFF Fisheries Laboratory has succeeded for the first time in measuring the long-term transport of 'northern component water' as this fast, dense, near-bottom stream descends into the deep ocean along the Continental Slope off East Greenland.

At the chosen location off Angmagssalik (Fig. 1) the MAFF current-meter array intercepts both the principal dense overflow from the Denmark Strait as well as a slower moving and less-dense derivative of more-distant overflows between Iceland and Scotland^{6,7}.

Following a nine-month trial of the mooring design, a total of 42 out of 44 current meters were recovered from two year-long deployments of our eight-mooring array between June 1987

and July 1989.

The distribution of mean current speeds for all 42 records (Fig. 2) confirms that the array was well placed to assess the transport of southward-directed flows denser than $\sigma_\theta = 27.80$. The uppermost instruments were close to that isopycnal, the deepest instruments lay within 20 m of the bed, and the presence of northward and eastward mean flows in the easternmost mooring of the array confirms that the overflowing waters were adequately closed off to the east. Data durations are adequate also, averaging 276 days.

The distribution of mean speeds was consistent on all deployments (Fig. 2): a cold, bottom-intensified flow parallel to the isobaths with mean speeds of up to $25\text{--}30 \text{ cm s}^{-1}$, maximum near-bottom current speeds of around 100 cm s^{-1} , a dominant fluctuation timescale of a few (1–10) days (compare to ref. 5), with amplitude similar to that of the mean, but with little sign of any systematic seasonal signal (compare to ref. 8) or indeed of any month-to-month variation greater than a few centimetres per second or a few degrees in direction.

Overlying the isopycnal distributions on the distribution of mean speeds provides a mean transport rate of 10.70 Sv for the deepening dense ($\sigma_\theta \geq 27.80$) flow passing south through this point on the East Greenland Continental Slope ($1 \text{ Sv} = 10^6 \text{ m}^3 \text{ s}^{-1}$). As mean flow speeds appear to be steady for records of this length—or indeed for any averaging period longer than 40–70 days—we expect this transport estimate to be stable also, but have a final array in the water at present to confirm this point (dots, Fig. 2a).

Of the 10.70 Sv passing through our section at $\sigma_\theta \geq 27.80$, a four-end member mixing model based on potential temperature, salinity, Freon-12 and silicate suggests that only 2.7 Sv originate in overflow through the Denmark Strait itself. The fourfold increase downstream is due to local entrainment of water as the fast, dense, near-bottom stream runs south along the Continental Slope, and to the presence on our section of dense water masses of other origins such as the water that overflows the Iceland–Scotland Ridge by the Faroe Bank Channel (plus its entrainment).

Several authors have advanced schemes for the circulation and exchange of water masses between the northern North Atlantic and the Norwegian–Greenland Sea. All such schemes in which the circulation is inferred rather than measured must

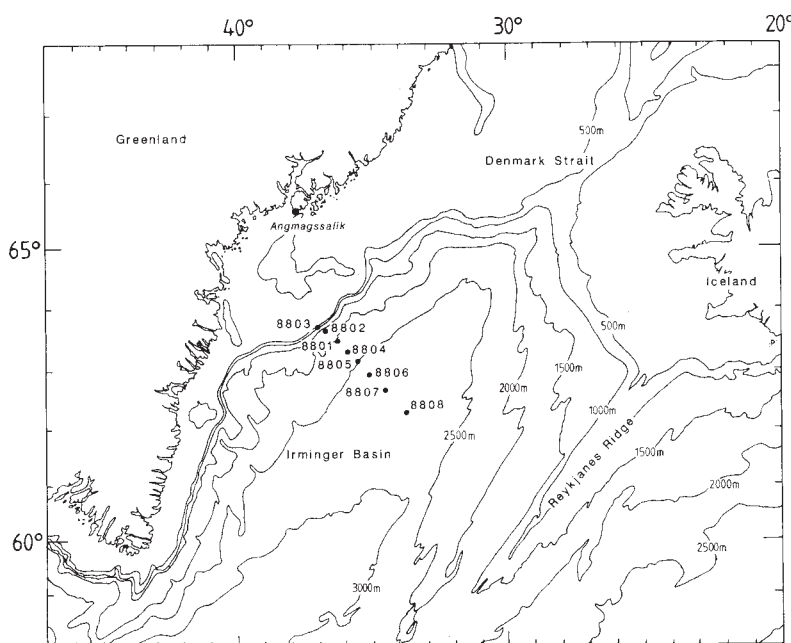


FIG. 1 Location of the eight-mooring current-meter array on the Continental Slope off East Greenland.

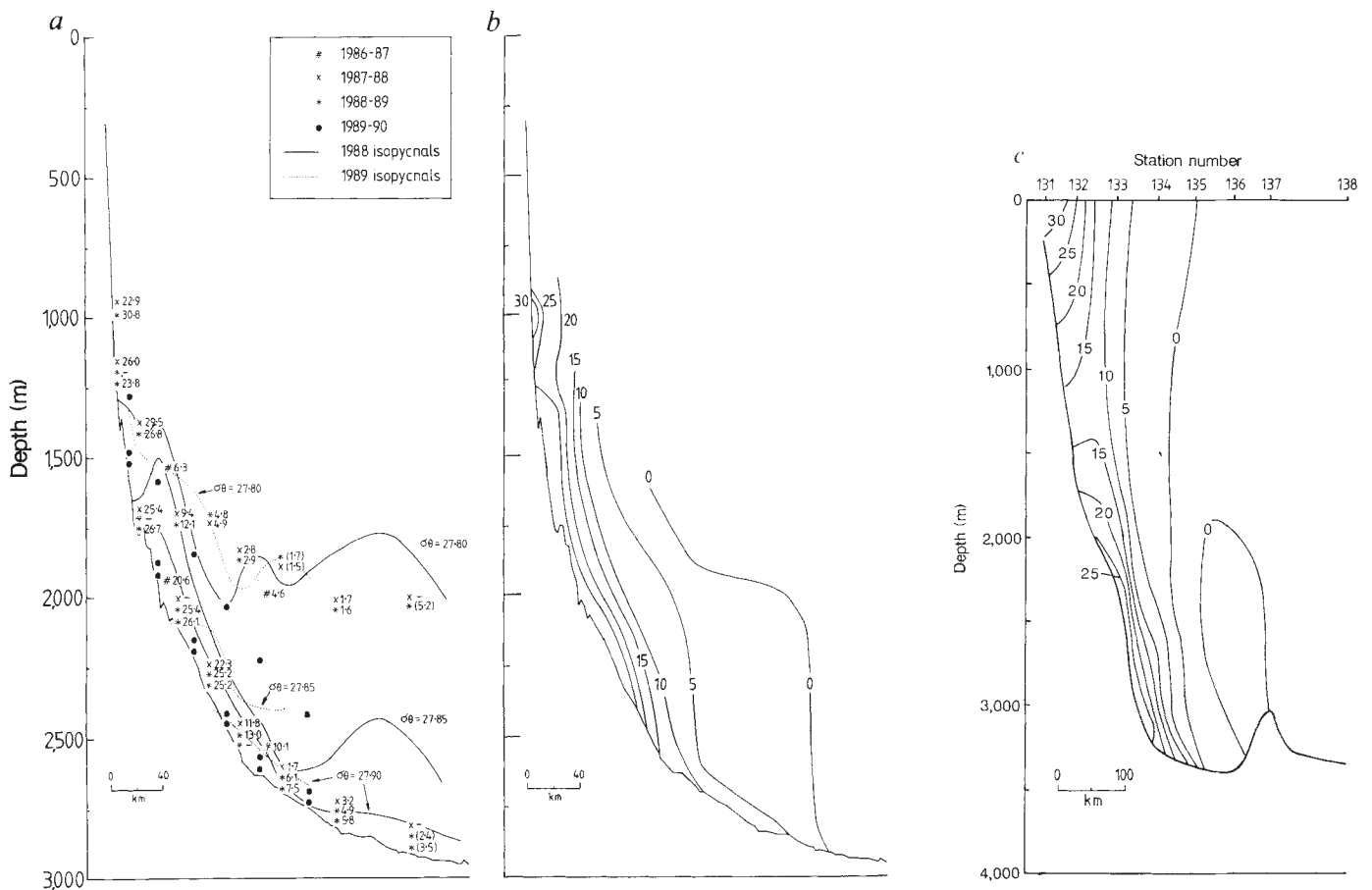
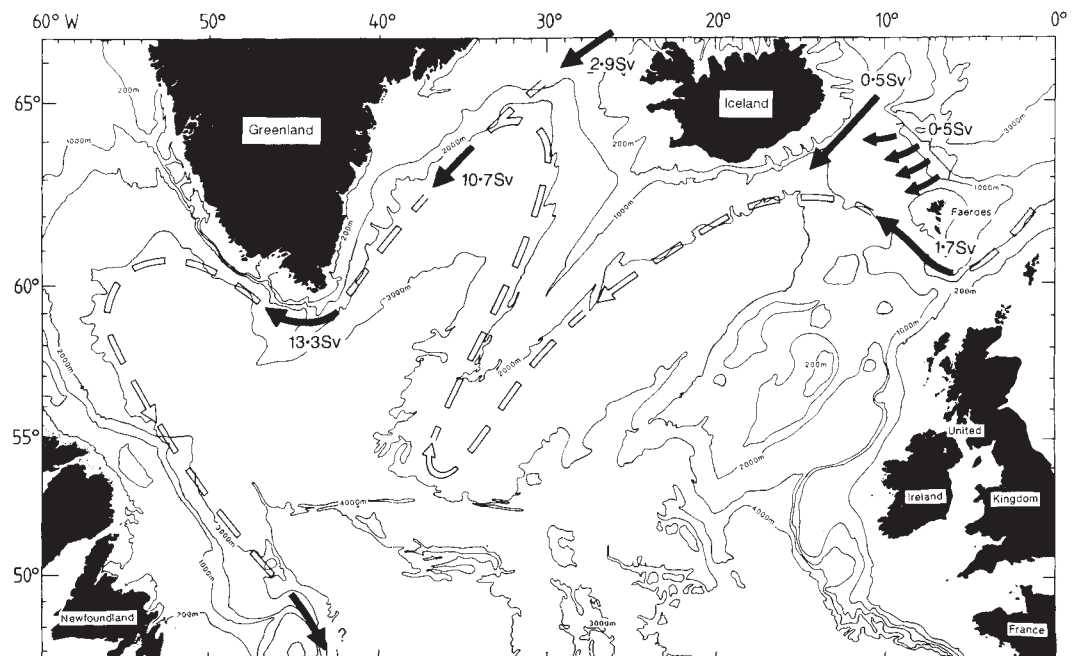


FIG. 2 a, Distribution of isopycnals ≥ 27.80 on the Angmagssalik array in July 1988 and July 1989 together with long-term mean-current speeds, all deployments (cm s^{-1}) with flows to the north and east in brackets. b,

Contours of long-term mean-current speed (cm s^{-1}) on the same section. c, Contours of geostrophic velocity south of Cape Farewell in 1978¹⁶.

FIG. 3 Transport scheme for waters denser than $\sigma_\theta = 27.80$ in the northern North Atlantic, based on all available measurements.



be more or less speculative as there are a number of ways in which our knowledge of high-latitude heat flux from ocean to atmosphere can be balanced by interbasin transports at prescribed temperatures.

So although Worthington^{9,10} and McCartney and Talley¹¹ achieve more or less comparable transports for the dense Deep Western Boundary Current (DWBC) as it flows south through 50° N (10 Sv and 11.1 Sv respectively), this similarity is based on radically different schemes; in Worthington's models, for example, the whole of this DWBC transport is composed of dense cold overflows across the Greenland-Scotland Ridge plus subsequent entrainment of warmer water, whereas in the model of McCartney and Talley three-quarters of the DWBC passing 50° N is produced through active formation of Labrador Sea water in the sub-polar basin. In another model¹² the DWBC is assumed to be composed solely of water of Iceland-Scotland origin, with the Denmark Strait outflow discounted.

We suggest that the body of direct measurements is now sufficiently complete and self-consistent to offer a worthwhile constraint on these models and their underlying assumptions. As summarized in Fig. 3, existing measurements suggest the following transport scheme for overflowing waters in the northern gyre denser than $\sim \sigma_\theta = 27.80$.

For the cold Denmark Strait overflow we find no reason to dispute the figure of 2.9 Sv published by Ross¹³. Our tracer model suggests almost exactly this figure and although Ross's current measurements are short (5 weeks), our own longer measurements suggest they may be adequate.

For the cold Faroe Bank Channel outflow, recent estimates^{8,14} suggest means in the range 1.5–1.9 Sv, with an additional 1.0 Sv overflowing through lesser notches in the Iceland–Faroe Ridge¹⁵.

Our Angmagssalik section will intercept much if not all of these overflows; we therefore assume that the stable transport of 10.7 Sv that we measure is made up of the sum of the overflows (~ 5.6 Sv) plus an equivalent (but undifferentiated) transport of warmer water entrained by these streams since their points of input.

Although we can reasonably assume that the process of entrainment is complete for the Iceland–Scotland overflow components, we cannot yet know how complete is entrainment by the Denmark Strait overflow. An early Canadian estimate¹⁶ for the deep flow rounding Cape Farewell suggests a transport of 13.3 Sv for $\sigma_\theta \geq 27.80$ (recalculated for our density range by R. A. Clarke, personal communication, 1989); though this estimate is based on geostrophic currents calculated from short-term hydrography, three current-meter moorings (six instruments) provided 60-day means for reference, and the resulting velocity distribution, reproduced here as Fig. 2c looks strikingly similar to the current we measure hugging the slope off Angmagssalik. We therefore believe the estimate is valid and that the progression in deep-water overflow and entrainment we describe, from 5.6 Sv at the points of overflow through 10.7 Sv off Angmagssalik to 13.3 Sv off Cape Farewell (and to ~ 15 Sv at 36° and 24° N? (refs 17, 18) may be close to a correct description for the production profile of Northern Component Water. This statement presumes no subsequent addition through direct deep- and bottom-water formation in the Labrador Sea itself, but we know of no evidence that such a mechanism exists. □

11. McCartney, M. S. & Talley, L. D. *J. phys. Oceanogr.* **14**, 922–935 (1984).
12. Pickart, R. S., Hogg, N. G. & Smethie, W. M. *J. phys. Oceanogr.* **19**, 940–951 (1989).
13. Ross, C. K. *Rapp. P.-v. Réun. Cons. perm. int. Explor. Mer.* **185**, 111–119 (1984).
14. Borenas, K. M. & Lundberg, P. A. *J. geophys. Res.* **93**, 1281–1292 (1988).
15. Meincke, J. in *Structure and Development of the Greenland–Scotland Ridge* (eds Bott, A., Savov, B., Talwani, C. & Thiede, D.), 635–650. (Plenum, New York, 1983).
16. Clarke, R. A. *Rapp. P.-v. Réun. Cons. perm. int. Explor. Mer.* **185**, 120–130 (1984).
17. Roemmich, D. J. *J. phys. Oceanogr.* **10**, 1972–1983 (1980).
18. Hall, M. M. & Bryden, H. L. *Deep Sea Res.* **29**, 339–359 (1982).

ACKNOWLEDGEMENTS. We acknowledge the assistance of scientists and crew of RV *Walther Herwig* in deploying our first trial moorings off Greenland in September 1986, and the help of K. Aagaard, Pacific Marine Environmental Laboratory, in providing both equipment and encouragement. A. Clarke, Bedford Institute of Oceanography, provided invaluable assistance.

The evolution of strontium isotopes in the upper continental crust

Frank McDermott & Chris Hawkesworth

Department of Earth Sciences, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK

RADIOGENIC isotopes in clastic sediments record how the average age and composition of the exposed continental crust has changed with time. Most of the continental crust is >1.5 Gyr old, and the upper crust is characterized by high Rb/Sr ratios. Critically, however, late Precambrian and younger sediments have surprisingly low depositional $^{87}\text{Sr}/^{86}\text{Sr}$ ratios, given their old source ages and high measured Rb/Sr (refs 1, 2). This paradox has prompted suggestions that the $^{87}\text{Sr}/^{86}\text{Sr}$ ratio in clastic sediments is buffered by a significant flux of new (low $^{87}\text{Sr}/^{86}\text{Sr}$) material from the upper mantle, with the implication that relatively large quantities of ^{87}Sr were generated in the sedimentary reservoir, and subsequently recycled into the upper mantle^{1,2}. We argue here that the strontium isotope paradox instead largely reflects the gradual increase in Rb/Sr with time in the upper continental crust. The Rb/Sr ratio of the upper crust is increased by intracrustal melting and fractional crystallization in plagioclase-bearing systems, and by weathering, erosion and sedimentation. The longer material resides in the continental crust, the more likely it is to be reworked by magmatic and sedimentary processes, and so to acquire high Rb/Sr ratios. One corollary is that $\sim 50\%$ less ^{87}Sr has been generated in clastic sediments in the past 2.0 Gyr than in models that assume constant crustal Rb/Sr (Fig. 3 of ref. 2).

The Sr-isotope evolution of the upper continental crust is determined by its age and Rb/Sr ratio, which in turn reflect the age and Rb/Sr ratio of material added from the upper mantle and the lower crust, and the rate of change of Rb/Sr in the upper crust. Rb/Sr ratios are increased by partial melting and fractional crystallization, and by weathering and erosion because Sr is more readily removed in solution to the oceans than Rb^{3,4}. Up to 90% of clastic sediments are derived from pre-existing sediments⁵, and this reprocessing of crustal material in the sedimentary system typically results in unusually high Rb/Sr ratios (>1) in clastic sediments. Thus, both magmatic and sedimentary processes in the continental crust tend to increase Rb/Sr, and so accelerate the rate of increase of $^{87}\text{Sr}/^{86}\text{Sr}$. Competing processes that reduce the rate of increase of $^{87}\text{Sr}/^{86}\text{Sr}$ include the introduction of material from the lower crust and upper mantle, and the preferential removal⁶ of ^{87}Sr to the mantle^{1,2}. But the observation that the measured Rb/Sr ratios of many clastic sediments are much higher than their time-integrated values is also consistent with a model in which secular increases in Rb/Sr dominate the Sr-isotope evolution of clastic sediments. In this model the present rate of increase in $^{87}\text{Sr}/^{86}\text{Sr}$ in clastic sediments is much higher than it was in earlier periods of Earth's history.

Time-integrated Rb/Sr ratios are those required to generate

Received 7 December 1989; accepted 16 March 1990.

1. Broecker, W. S. & Denton, G. H. *Geochim. cosmochim. Acta* **53**, 2465–2501 (1989).
2. NEA/OECD Review of the continued suitability of the dumping site for radioactive waste in the North Atlantic (Paris, 1985).
3. Broecker, W. S. & Peng, T. H. *Tracers in the sea*, 317 (Eldigio, Palisades, 1982).
4. Warren, B. A. in *Evolution of Physical Oceanography* (eds Warren, B. A. & Wunsch, C.) 6–41 (MIT Press, Cambridge, 1981).
5. Worthington, L. V. *Deep Sea Res. Suppl.* **16**, 421–432 (1969).
6. Swift, J. H. *Deep Sea Res.* **11**, 1339–1355 (1984).
7. Livingston, H. D., Swift, J. H. & Ostlund, H. G. *J. geophys. Res.* **90**, 6971–6982 (1985).
8. Saunders, P. M. *J. phys. Oceanogr.* **20**, 29–43 (1990).
9. Worthington, L. V. *Deep Sea Res.* **17**, 77–84 (1970).
10. Worthington, L. V. *The Johns Hopkins Oceanography Studies* **6**, 110 pp. (1976).

the present-day $^{87}\text{Sr}/^{86}\text{Sr}$ ratios in a sample since it, or its crustal precursors, were derived from the upper mantle, as indicated by its model Sm-Nd age. They provide a measure of increases in Rb/Sr due to intracrustal processes, and the difference between time-integrated and measured Rb/Sr ratios reflect previous observations that Rb-Sr model ages are typically less than Sm-Nd model ages in fine-grained clastic sediments^{7,8}. Figure 1 illustrates the time-integrated and measured Rb/Sr ratios for selected sediment suites. The two may be similar either in rocks that seem to have formed with little fractionation of Rb/Sr, as in New Mexico⁹ (Fig. 1d), or in those for which the time elapsed since deposition is very much greater than the age of their source terranes at the time of deposition (such as Isua¹⁰, Fig. 1a). However, in the majority of clastic sediments the measured Rb/Sr ratios are significantly higher than the time-integrated values. In the simplest interpretation, the average Rb/Sr ratios of modern clastic sediments are higher than in ancient clastic sediments and models for the evolution of the continental crust need to take this into account.

If Rb/Sr increases with time in the clastic sedimentary system, the depositional $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of clastic sediments should tend to increase with the crustal residence age of their source rocks at the time of deposition. Figure 2 summarizes the depositional $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of clastic sediments^{7-9,11} and river particulates¹² plotted against the difference between their model Sm-Nd and their stratigraphic ages. This age difference represents the crustal residence age at the time of deposition, as the Sm-Nd model ages are taken to indicate when the sediment source rocks, or their crustal precursors, were derived from the upper mantle. The data are scattered, but overall the higher depositional

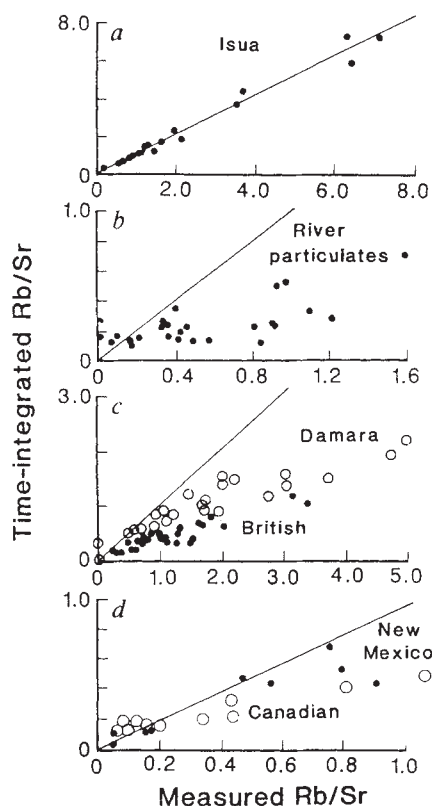


FIG. 1 A plot of measured Rb/Sr ratios against time-integrated Rb/Sr ratios for clastic sediments⁷⁻¹¹ and river particulates¹². Note that many of the Isua samples with unusually high Rb/Sr are from a 'boulder bed' of uncertain origin¹⁰. With the exception of the Isua¹⁰ and New Mexico⁹ sediments, time-integrated Rb/Sr ratios are typically lower than measured ratios which largely reflects intracrustal Rb/Sr fractionation (for discussion, see text).

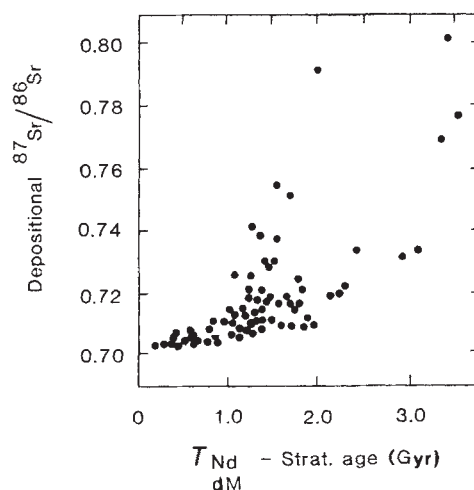


FIG. 2 Depositional $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of post-Archaean fine-grained clastic sediments^{7-9,11} and river particulates¹² plotted against the difference between their Sm-Nd model α and stratigraphic age (relative to depleted mantle) age. This age difference represents the crustal residence age of their source rocks at the time of deposition.

$^{87}\text{Sr}/^{86}\text{Sr}$ values are clearly associated with source regions that have the longer crustal residence times at the ages of sedimentation.

In detail, sediments derived from young crustal regions ($T_{\text{dm}}^{\text{Nd}} - \text{Strat. age} \sim 0$) have depositional $^{87}\text{Sr}/^{86}\text{Sr} < 0.705$ (Fig. 2), and as the source regions get older the sediments have higher depositional $^{87}\text{Sr}/^{86}\text{Sr}$. If Rb/Sr was constant in the source of clastic sediments, depositional $^{87}\text{Sr}/^{86}\text{Sr}$ ratios would be proportional to the crustal residence ages of the source rocks, and sample suites would plot on straight lines in Fig. 2. But the concave-up nature of the available data arrays indicates that those sediment sources with the longest residence ages have the highest time-integrated Rb/Sr ratios, as reflected in their $^{87}\text{Sr}/^{86}\text{Sr}$ ratios at the time of sedimentation. As outlined above, this is interpreted in terms of an accumulative increase in Rb/Sr due to prolonged crustal reworking by both magmatic and sedimentary processes. The scatter of data, for example between samples that have very different $^{87}\text{Sr}/^{86}\text{Sr}$ ratios at similar $T_{\text{dm}}^{\text{Nd}} - \text{Strat. age}$ (Fig. 2), may be due to differences in the relative contributions from magmatic and sedimentary processes, or perhaps more likely, to differences in the frequency with which material was reworked in different crustal areas.

The key issue, therefore, is how to explain the relatively low depositional $^{87}\text{Sr}/^{86}\text{Sr}$ ratios (< 0.725), and hence low inferred time-integrated Rb/Sr, of the majority of Phanerozoic to Recent clastic sediments, given their high measured Rb/Sr ratios (typically > 1) and old average Sm-Nd model ages (> 1.5 Gyr). Either the Rb/Sr values of upper crustal source rocks were lower in the past, and/or the depositional $^{87}\text{Sr}/^{86}\text{Sr}$ values are too low, and so some mechanism must be found to reduce them. In practice both effects are likely to have contributed, but the more the latter is invoked the greater the required flux of new material from the mantle to the clastic sedimentary system.

The depositional $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of post-Archaean sediments for which Nd isotope data are also available are plotted against time in Fig. 3. The majority are < 0.725 , even though their Sm-Nd model ages are typically in the range 1.5–2.0 Gyr. The straight lines are the evolution paths for sediments assuming that their Rb/Sr ratios were constant and similar to that of average post-Archaean shale (0.8) (ref. 13), and that their crustal source rocks were derived from the upper mantle 1.5–2.0 Gyr ago. Clearly, the calculated $^{87}\text{Sr}/^{86}\text{Sr}$ ratios are much higher than those observed in late Precambrian and younger sediments, and it was

this discrepancy that prompted suggestions that $^{87}\text{Sr}/^{86}\text{Sr}$ in the sedimentary reservoir was buffered externally^{1,2}. There is little evidence that it could be buffered significantly by contributions from the lower crust, and so it was argued that it is more likely to have been buffered by the introduction of new material from the upper mantle^{1,2}.

Crustal growth models suggest that ~10% of the continental crust was generated in the last 1 Gyr¹⁴⁻¹⁶. Clastic sediments that had a typical $^{87}\text{Sr}/^{86}\text{Sr}=0.71$ at 1 Gyr (Fig. 3), would have $^{87}\text{Sr}/^{86}\text{Sr}=0.743$ at the present day if their average $\text{Rb}/\text{Sr}=0.8$ had remained constant over that period. This contrasts sharply with the average present-day $^{87}\text{Sr}/^{86}\text{Sr}\sim 0.716$ in clastic sediments, presumably due to the combined effects of increasing Rb/Sr in their upper crustal sources, and the introduction of low $^{87}\text{Sr}/^{86}\text{Sr}$ new crust from the mantle. If the difference between 0.743 and 0.716 was solely due to the introduction of mantle material to a continental sedimentary reservoir of 1.88×10^{24} g (ref. 13) containing 200 p.p.m. Sr, then $\sim 7 \times 10^{20}$ g of mantle Sr are required. Assuming that new crust contains 400 p.p.m. Sr (ref. 13), this represents ~70% of the Sr thought to have been derived from the upper mantle in the past 1 Gyr, and it is therefore unrealistically high.

The sedimentary record is biased towards relatively young crust because young rocks are eroded preferentially. This bias may be expressed in an erosion coefficient K (ref. 17), and the best estimates are for an average $K=2$; that is, the proportion of juvenile (<0.5 Gyr) to old (for example, ~2.0 Gyr) crust contained in a clastic sediment is typically ~80% higher than the proportion in crust exposed to weathering¹⁷. The bias towards new crust will be less over the past 1 Gyr because relatively less new crust has been generated, and $K=1.6$ is consistent with the difference between the average Sm-Nd model ages of recent sediments and the average age of the upper crust¹⁸. If new crustal material is partitioned between the sedimentary reservoir and the rest of the continental crust in proportion to their present-day masses, then ~8% of the new crust generated

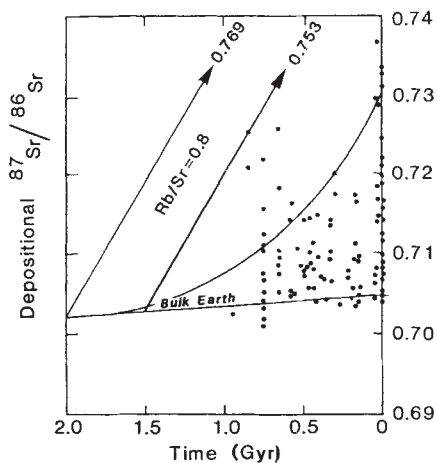


FIG. 3 Depositional $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of post-Archaeon sediments for which Nd isotope data are available^{7-9,11,12} plotted against their stratigraphic age. Most have $^{87}\text{Sr}/^{86}\text{Sr}$ ratios lower than 0.725, in contrast with the much higher depositional $^{87}\text{Sr}/^{86}\text{Sr}$ ratios predicted by their present-day high Rb/Sr (average=0.8) and relatively old (1.5–2.0 Gyr) Sm-Nd model ages. Thus, the straight-line evolution paths from a initial value of 0.12 at 1.8 Gyr, to 0.8 at the present day. Such increases in Rb/Sr are attributed to partial melting and to erosion and sedimentation, and together with the addition of a relatively small contribution of low $^{87}\text{Sr}/^{86}\text{Sr}$ mantle-derived material, they can account for the low depositional $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of many post-Archaeon clastic sediments.

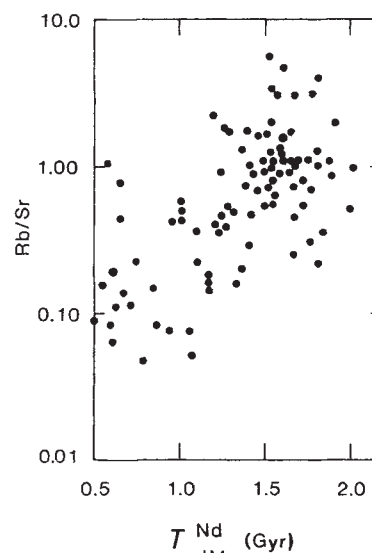


FIG. 4 Rb/Sr ratios of post-Archaeon clastic sediments for which Nd isotope data are also available^{7-9,11} are plotted against Sm-Nd model ages. Rb/Sr ratios are variable, but show a broad increase with increasing Sm-Nd model ages.

in the past 1 Gyr (1.89×10^{23} g) would have been added to the sedimentary reservoir. $K=1.6$ increases that to 12% (2.88×10^{23} g), but assuming that the new crust contained 400 p.p.m. Sr (ref. 13) it only reduces the present-day $^{87}\text{Sr}/^{86}\text{Sr}$ in clastic sediments from 0.743 to 0.734. This is just 33% of the shift required to produce the observed average of 0.716, and so the introduction of mantle Sr does not seem to be the dominant effect in determining the evolution of $^{87}\text{Sr}/^{86}\text{Sr}$ in the sedimentary reservoir.

Our preferred interpretation is that the continuing processes of intracrustal melting and of weathering and erosion have increased Rb/Sr in the upper crustal sources of clastic sediments. Figure 4 indicates that sediments with older model ages tend to have higher Rb/Sr ratios. In part, that may reflect the lower Rb/Sr ratios of younger new crust¹⁹, but clastic sediments still have Rb/Sr ratios which are typically 10 times those in new crustal material. The longer material resides in the upper crust, the more likely it is to be affected by intracrustal reworking processes, and so to have high Rb/Sr ratios (Figs 2 and 4). Such secular increases in Rb/Sr may then be incorporated into an average evolution curve for Sr isotopes in post-Archaeon clastic sediments (Fig. 3). The evolution path chosen for clastic sediments illustrates the effect of increasing Rb/Sr from an initial value of 0.12 when this crust was generated at 1.8 Gyr, to the average present-day value of 0.8 for post-Archaeon shale¹³. As illustrated in Fig. 3, such a model of increasing Rb/Sr with time in the clastic sedimentary reservoir can readily account for the relatively low $^{87}\text{Sr}/^{86}\text{Sr}$ ratios in late Precambrian and younger sediments. For example, the calculated $^{87}\text{Sr}/^{86}\text{Sr}$ at 0.5 Gyr is 0.7185, and together with the likely mantle contributions discussed above (1.15×10^{20} g Sr), is consistent with the estimated average $^{87}\text{Sr}/^{86}\text{Sr}=0.716$ in Phanerozoic clastic sediments. Although these calculations are necessarily rough, the available evidence indicates that Rb/Sr fractionation in the upper crust is the dominant effect that determined the evolution of Sr isotopes in the sedimentary reservoir over the past 1 Gyr.

Received 17 January; accepted 13 March 1990.

- Goldstein, S. L. in *Workshop on the growth of continental crust* (ed. Ashai, L. D.) LPI Tech. Rpt. 88-02, 65–67 (Lunar and Planetary Institute, Houston, 1988).
- Goldstein, S. L. *Nature* **336**, 733–738 (1988).
- Dasch, E. J. *Geochim. cosmochim. Acta* **33**, 1521–1556 (1969).
- Martin, J. M. & Meybeck, M. *Mar. Chem.* **7**, 173–206 (1979).
- Veizer, J. & Jansen, S. L. *J. Geol.* **93**, 625–643 (1985).

6. Albarède, F., Michard, A., Minster, S. F. & Michard, G. *Earth planet. Sci. Lett.* **55**, 229–236 (1981).
7. Davies, G. R., Gledhill, A. R. & Hawkesworth, C. J. *Earth planet. Sci. Lett.* **75**, 1–12 (1985).
8. McDermott, F., Harris, N. B. W. & Hawkesworth, C. J. *Tectonophysics* **161**, 257–270 (1989).
9. Nelson, B. K. & DePaolo, D. J. *J. sed. Petrol.* **58**, 348–357 (1988).
10. Jacobsen, S. B. & Dyrme, R. F. *J. geophys. Res.* **93**, 338–354 (1988).
11. Samson, S. D., McClelland, W. C., Patchett, P. J., Gehrels, G. E. & Anderson, R. G. *Nature* **337**, 705–709 (1989).
12. Goldstein, S. J. & Jacobsen, S. B. *Earth planet. Sci. Lett.* **87**, 249–265 (1988).
13. Taylor, S. R. & McLennan, S. M. *The continental crust: its composition and evolution*, 312 (Blackwell Scientific, London, 1985).
14. McLennan, S. M. & Taylor, S. R. *J. Geol.* **90**, 347–361 (1982).
15. Patchett, P. J. & Arndt, N. T. *Earth planet. Sci. Lett.* **78**, 329–333 (1986).
16. Allègre, C. J. *Tectonophysics* **81**, 109–132 (1982).
17. Allègre, C. J. & Rousseau, D. *Earth planet. Sci. Lett.* **67**, 19–34 (1984).
18. Goldstein, S. L., O'Nions, R. K. & Hamilton, P. J. *Earth planet. Sci. Lett.* **70**, 221–236 (1984).
19. Ellam, R. M. & Hawkesworth, C. J. *Geology* **16**, 314–317 (1988).

ACKNOWLEDGEMENTS. We thank R. Ellam and N. Harris for many discussions on the problems of constraining models for the evolution of the continental crust, and S. Goldstein and P. Taylor for their reviews.

Sediment-induced amplification and the collapse of the Nimitz Freeway

S. E. Hough*, P. A. Friberg*, R. Busby*, E. F. Field*, K. H. Jacob* & R. D. Borchardt†

* Lamont Doherty Geological Observatory of Columbia University, Palisades, New York 10964, USA

† United States Geological Survey, 345 Middlefield Road, Menlo Park, California 94025, USA

THE amplification of ground motion by low-seismic-velocity surface sediments is an important factor in determining the seismic hazard specific to a given site. The $M_s = 7.1$ Loma Prieta earthquake of 17 October 1989 was the largest event in the contiguous United States in 37 years, and yielded an unparalleled volume of seismic data from the main shock and aftershock sequence¹. These data can be used to image the seismic source, to study detailed Earth structure, and to study the propagation of seismic waves both through bedrock at depth and through sediment layers near the surface. Near the edge of San Francisco Bay, site conditions vary considerably on scales of hundreds of metres. The collapsed section of the two-tiered Nimitz Freeway in Oakland was built on San Francisco Bay mud, whereas stiffer alluvial sediments underlie a southern section that was damaged but did not collapse. Here we analyse high-quality, digital aftershock recordings from several sites near the Nimitz Freeway, and conclude that soil conditions and resulting ground-motion amplification may have contributed significantly to the failure of the structure.

Seismologists have long known that repeats of major earthquakes in California are inevitable. Despite substantial property damage resulting from the Loma Prieta earthquake, the large number of undamaged buildings and the low loss of life indicate the success of earthquake preparedness in the region. To the general public, however, the highly visible effects of this earthquake are likely to have a lasting impact.

Here we examine the role of site effects in the collapse of a 1.4-km stretch of the Nimitz Freeway (Interstate 880) in West Oakland, one of the most tragic consequences of the Loma Prieta earthquake. Engineers had identified design weaknesses and proposed retrofitting the structure; the issue was discussed extensively in news magazines and papers in the weeks following the earthquake. Clearly, structural design played a large part in the freeway collapse. However, the southern half of the two-tiered freeway, known as the Cypress Structure, which is of virtually identical design, was damaged but did not collapse (Fig. 1).

Evidence from eyewitnesses suggests that triggered collapse of adjacent segments may have occurred. In two locations, one

in the northern half of the structure over a railroad line and one at point B in Fig. 1, short sections of the freeway were built with three-pillar support of the bottom tier instead of the two-pillar support used elsewhere. Over the railroad line, the supports are also oblique, and the second tier remained standing between the two three-legged supports. At point B, the collapse ends; the supports further south were damaged, but did not fall.

Figure 1 shows surficial geology. The northern segment of the freeway is built on San Francisco Bay mud/fill, which characterizes the land adjacent to the west. The amplification of ground motion by surficial layers has been discussed in literature^{2–5}, dating back to Milne (1898). Wood observed that high levels of damage were correlated to near-surface sediments during the 1906 San Francisco earthquake⁴. More recently, sediment amplification of ground motion contributed to high levels of damage and loss of life in down-town Mexico City, some 300 km from the 1985 Michoacan, Mexico earthquake^{5,6}.

Seismic shear-wave velocities in San Francisco Bay mud have been estimated at 90–130 m s⁻¹ (ref. 6), and their potential to contribute to seismic hazard through sediment-induced amplification has been well documented^{7–9}. The southern end of the Cypress Structure is built on Quaternary sediments, with seismic shear-wave velocities estimated between 200 and 400 m s⁻¹ and correspondingly less severe amplification⁷.

Working under the auspices of the National Center for Earthquake Engineering Research (NCEER) and the Incorporated Research Institution for Seismology (IRIS), six stations were occupied in the vicinity of the Cypress Structure (Fig. 1). These sites included a mud site (~600 m west of the collapsed freeway), an alluvium site (~250 m east of the uncollapsed freeway), and a reference station on Franciscan rocks, ~3 km away. Sensors were also installed on the top of the uncollapsed second tier for 15 h to study the structural response of the freeway. During this time, a number of unidentified, high-amplitude noise events were recorded.

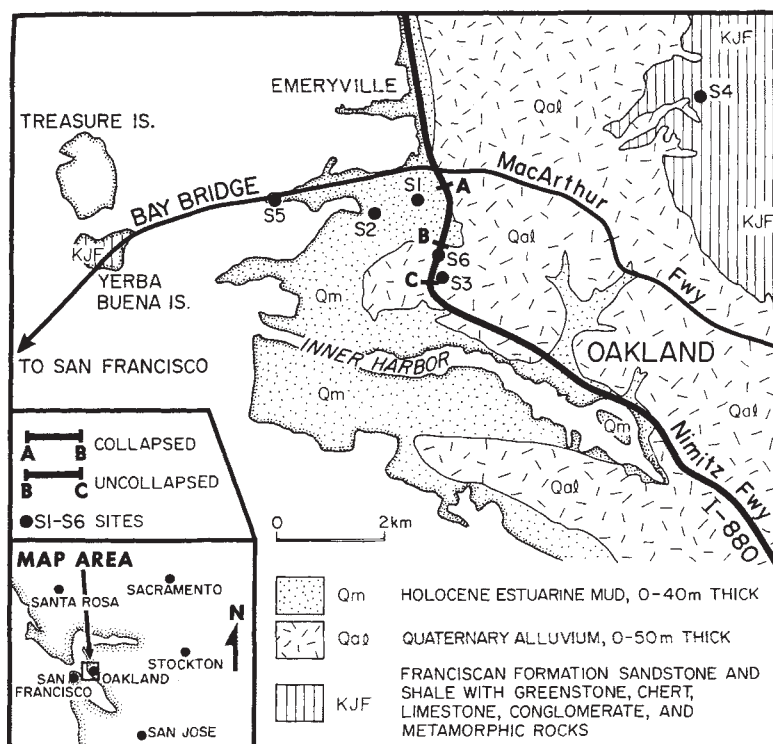
Stations S1–S5 (Fig. 1) were equipped with 2-Hz sensors and digital recorders. At three sites (S2, S3 and S4), the seismometers were deployed on concrete surfaces or on linoleum over concrete. At S1 and S5, the sensors were buried in soil to reduce noise at windy sites. Recording was at 100 samples per second. Triggering occurred when short-term amplitudes exceeded long-term signal levels by a factor of >4.5. Instrument response is approximately flat to velocity between 2 and 40 Hz. Because the data are used to compute spectral ratios, no instrument correction is performed. Five-second sensors were used for recording at station S6, on the freeway. Sixteen magnitude 2.4–4.4 aftershocks were recorded on at least three of five of the ground stations S1–S5. Magnitudes are estimated with data from CALNET, a seismic network operated by the United States Geological Survey.

Figure 2a shows ground motion from a magnitude 4.1 aftershock recorded on one of the horizontal components (NS) on mud, alluvium and rock. Figure 2b shows ground motion from the vertical components.

A fast Fourier transform is applied to a 10-s window bracketing the highest-amplitude shear arrivals, with smoothing over ± 0.3 Hz. Figure 3a presents spectral ratios for mud/rock sites using the time series shown in Fig. 2a as well as from a magnitude 4.4 aftershock. As expected, the response of San Francisco Bay muds yields sharply peaked amplification relative to rock sites.

A sediment layer is expected to give rise to resonances with frequencies given by the relationship $f = \beta/4H$, where β is the shear velocity, H is the layer thickness, and f is the resonance frequency of the fundamental mode. Differences in the amplitudes of the peaks near 1.5 and 3 Hz result from the variance of spectral estimates and the increase in variance that results from the process of taking ratios. Thus, the confidences of the amplitude of the sharpest peaks in Fig. 3 are low. A reduction in variance would necessitate a reduction in resolution of spectral peaks; a rigorous analysis is beyond the scope of this report.

FIG. 1 Map of Oakland showing near-surface geology¹¹. 'Bay mud' indicates areas partly covered with fill. Stations S1-S6 are sites occupied during this study. AC indicates extent of the two-tiered section of the Nimitz Freeway, known as the Cypress structure. AB indicates the segment that collapsed; BC indicates the segment that was damaged but did not collapse.



We note, however, that amplification is a factor ≥ 5 for frequencies between 1 and 4 Hz. Resonant frequencies in this range imply sediment thickness of 5–30 m. The thickness of the mud layer is constrained in some places from geotechnical surveys, and is known to vary considerably, with a maximum observed thickness of 40 m (ref. 8). Our estimates are thus consistent with expectations; further analysis of the observed resonances may be useful to constrain the depth of the mud layer at sites where geotechnical information does not exist.

For the purpose of understanding the role of site response in the Nimitz Freeway collapse, the pertinent comparison is between ground motion on mud and on alluvium. Figure 3b presents a spectral ratio of mud/alluvium. The reduced promi-

nence of the 1.5-Hz peak suggests that this response is common to both S1 and S3 and represents a resonance of the alluvium layer, although ground motions on mud are still amplified by a factor of 2–3 at 1.5 Hz. Relative to the alluvium site, horizontal ground motions on the mud site were amplified by factors of 5–8 at frequencies between 2 and 5 Hz. Ground motion recorded on the vertical components is of significantly lower amplitude than on the horizontal components and is not as dramatically amplified.

The response of a two-tiered freeway is complex, involving the superposition of many modes. Evidence that the ground motion and structural frequencies coincided comes from the almost-perfect correlation between near-surface sediments and

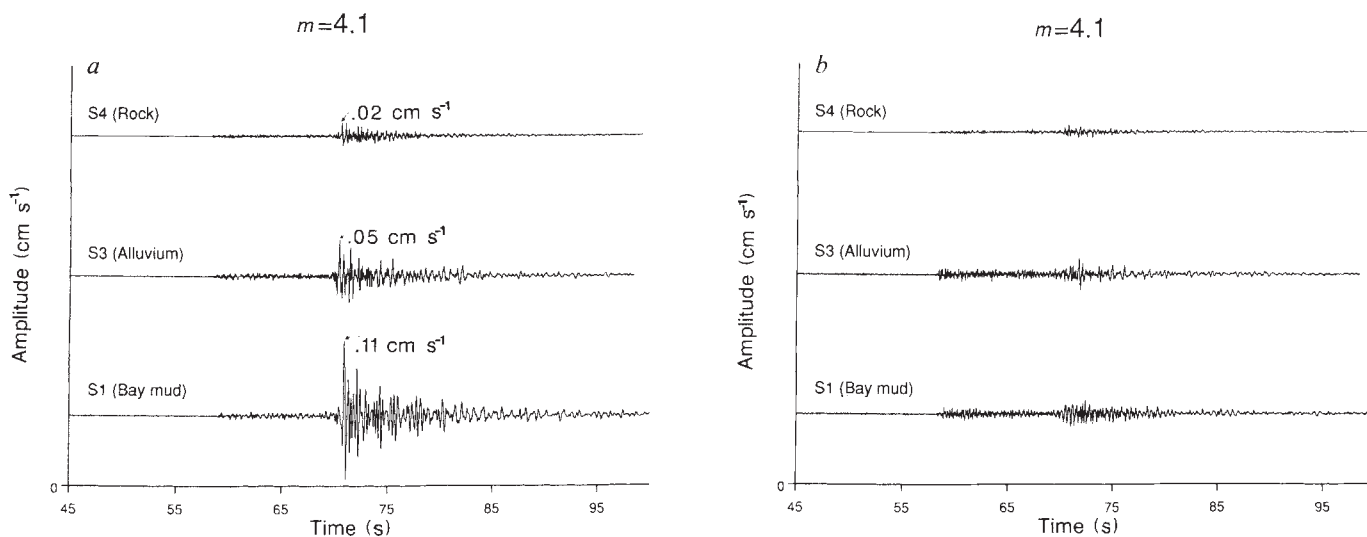


FIG. 2 *a*, Seismograms (amplitudes proportional to velocity) recorded on the north-south horizontal components for a magnitude 4.1 aftershock at S1 (on mud near the collapsed freeway), S3 (on alluvium near the uncollapsed freeway), and at S4 (on Franciscan rock in the Oakland Hills). Data are not corrected for instrument response, which is essentially flat to velocity

between 2 and 40 Hz. All three traces are plotted at the same scale. Peak velocities are shown for each trace. *b*, Same event, axes and scale as *a*, for vertical component of motion. Peak velocities are 0.019 cm s^{-1} (mud), 0.024 cm s^{-1} (alluvium) and 0.013 cm s^{-1} (rock).

collapse illustrated in Fig. 1. To make preliminary observations of the resonant periods of the structure, we installed instruments on the uncollapsed second tier of the structure, at station S6 in Fig. 1. Although these data gave some information about structural response, a definitive study requires more instrumentation than the single set of seismometers available to us at the time. A later study of structural response was made by engineers at the University of California, Berkeley¹⁰, who made simultaneous recordings of ambient noise at both levels of the freeway and on the ground. Using the simultaneous surface recordings to separate source and structural effects, this study found resonance frequencies of various horizontal modes between 1.6 and 4.5 Hz, with a fundamental transverse mode at 2.5 Hz. Because these modes coincide almost perfectly with the frequencies at which ground motions on mud are most strongly amplified relative to ground motions on alluvium, the ground motion experienced during the main shock would have efficiently excited structural resonances. This type of match between excitation and structural frequencies will contribute greatly to structural damage.

In this study, preliminary observations show significant amplification of horizontal ground motion on San Francisco

Bay mud adjacent to the collapsed section of the Cypress structure relative to motion on alluvium sites near the uncollapsed section of the Cypress structure relative to motion on alluvium sites near the uncollapsed section. The mud site was 600 m from the freeway collapse, and sediment thickness is likely to vary between site S1 and the freeway. Assuming, however, that observations at S1 are characteristic of site response on mud in the vicinity, our observations suggest that these sediment resonances could have had a significant role in contributing to freeway damage.

Identifying areas where strong site effects will occur is easy when damage has already occurred. Although forensic studies are valuable to document the role of ground-motion amplification, the phenomenon is already relatively well understood from a theoretical and observational perspective. We suggest that future research efforts should have a two-fold focus: (1) to identify sediment sites where known or suspected impedance contrasts suggest that site effects may occur in future earthquakes; and (2) specific studies where sediment-induced amplification has been documented but where proposed development necessitates in-depth investigations. □

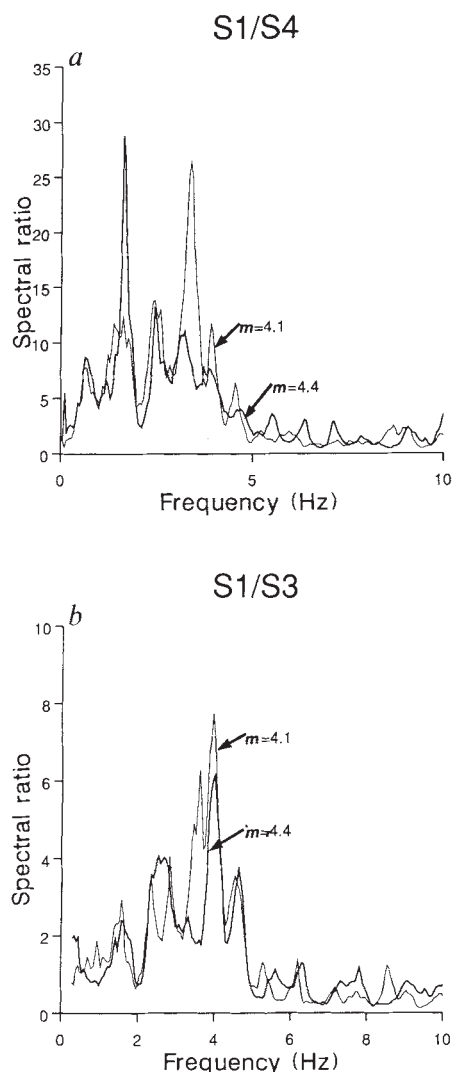


FIG. 3 a, Spectral ratios using stations S1/S4 (mud/rock sites) for the event used on Fig. 2a as well as for the largest recorded aftershock (magnitude 4.4). Although the amplitude of the sharpest peaks differs between the two events, we note that the confidence of these peaks is poor, and that amplification is at least a factor of 5 between 1 and 4 Hz. b, Spectral ratios using stations S1/S3 (mud/alluvium sites) are shown for the same events as used in a.

Received 7 December 1989; accepted 26 February 1990.

1. Earthquake Engineering Research Institute, Preliminary Reconnaissance Report, Loma Prieta Earthquake, October 17, 1989 EERI89-03, 51 pp. (1989).
2. Borchardt, R. D. *Bull. seism. Soc. Am.* **60**, 29–61 (1970).
3. Milne, J. *Seismology*, 1st edn 320 pp. (Kegan Paul, Trench, Truber, London, 1898).
4. Wood, H. O. *Carnegie Inst. Washington Publ.* **87**, 220–245 (1908).
5. Singh, S. K. *et al. Earthquake Spectra* **4**, 653–673 (1988).
6. Lermo, J., Rodriguez, M. & Singh, S. K. *Earthquake Spectra* **4**, 805–814 (1988).
7. Borchardt, R. D. *U.S. Geol. Surv. Prof. Paper* 941-A (1975).
8. Borchardt, R. D. & Gibbs, J. F. *Bull. seism. Soc. Am.* **66**, 467–500 (1976).
9. Rogers, A. M., Tinsley, J. C. & Borchardt, R. D. *U.S. Geol. Surv. Prof. Paper* 1360, 221–248 (1985).
10. Nims, D. K., Miranda, E., Aiken, I. D., Whittaker, A. S. & Bertero, V. V. *Rep. UCB/EERC-89/16* (1989).
11. Borchardt, R. D., Gibbs, J. F. & Lajoie, K. R. *U.S. Geol. Surv. Map MF 709* (1975).

ACKNOWLEDGEMENTS. The L-DGO portion of this research was supported by NCEER, and field deployment was supported by IRIS/PASSCAL and NCEER, with funding from the National Science Foundation. We acknowledge cooperation from Caltrans, and the Lamont authors gratefully acknowledge the cooperation and hospitality of the United States Geological Survey in Menlo Park. We thank N. Seeber, M. Anders and W. Joyner for constructive reviews of the manuscript. Lamont-Doherty Geological Observatory contribution 4604.

Fault reactivation in the central Indian Ocean and the rheology of oceanic lithosphere

Jonathan M. Bull & Roger A. Scrutton

Department of Geology and Geophysics, West Mains Road, Edinburgh EH9 3JW, UK

THE intraplate deformation in the central Indian Ocean basin is a well-known example of a deviation from an axiom of plate tectonics: that of rigid plates with deformation concentrated at plate boundaries. Here we present multichannel seismic reflection profiles which show that high-angle reverse faults in the sediments of the central Indian Ocean extend through the crust and possibly into the uppermost mantle. The dip of these faults, which we believe result from the reactivation of pre-existing faults formed at the spreading centre, is $\sim 40^\circ$ in the basement, which is consistent with the distribution and focal mechanisms of earthquakes on faults now forming at spreading centres. This style of deformation, coupled with the observation of large earthquakes in the mantle lithosphere, indicates that brittle failure of the oceanic lithosphere may nucleate in the vicinity of the brittle/ductile transition and propagate through the crust.

The intraplate deformation occurs in lithosphere of age 60–90 Myr that formed at a fast spreading rate (6 cm yr⁻¹) at the southeast Indian Ocean ridge. It manifests itself as a diffuse

zone of compressional and strike-slip earthquakes¹, high localized heat-flow², geoid anomalies³ and tectonic deformation^{4,5}. The tectonic deformation can be considered to be occurring on two spatial scales: the first order is represented by long-wavelength (100–300 km), large-amplitude (1–2 km) undulations of oceanic basement and overlying sediments; and the second order by unusual high-angle reverse faults and associated folds in the basement cover. There has been much speculation as to whether the first order of deformation represents buckling^{5,6}, inverse boudinage⁷ or some form of block faulting⁸. The high-angle reverse faults have been less well studied and are the subject of this paper.

The high-angle faults were first documented by Eittreim and Ewing⁴ and then further studied by Weissel *et al.*⁵ and Geller *et al.*². They found that they offset both basement and overlying sediments, dipped at angles $>65^\circ$ in the sedimentary cover and had a strike of 100° E. Weissel *et al.*⁵ suggested that they arose from the reactivation of the pre-existing structural fabric in basement. Recently, results from ODP Leg 116 (ref. 9) have shown that this reactivation started 7 Myr BP and Curray and Munasinghe¹⁰ have shown that the onset of deformation can be correlated over a wide area. Here we discuss results from RRS *Charles Darwin* Cruise 28, during which the first multichannel seismic reflection profiles were collected over the intraplate deformation (Fig. 1). These results not only confirm these pre-

vious observations but add important new findings about the character, origin and mode of reactivation of the faults.

The *Darwin* seismic reflection profiles clearly show that the vast majority of the high-angle faults are reverse in nature and few are normal, a conclusion which could not be made from the profiles used by Weissel *et al.*⁵. The reverse faults can be divided into those downthrowing to the north, and those downthrowing to the south. In the survey area approximately equal numbers of faults downthrow in each direction. The new multichannel profiles image (Fig. 2) the faults in basement that dip to the north, their dip being $\sim 35\text{--}40^\circ$ at this depth. In some instances the faults are resolved down to 10-s two-way travel time, roughly the expected level of the oceanic Moho^{8,11}. The faults, which appear slightly listric on the time section, are approximately planar after depth conversion. There are additional north-dipping events at lower angles that do not intersect the basement/cover interface that could be other, little reactivated, faults. Faults dipping to the south, on the other hand, are not well resolved. This non-resolution, together with the fact that the offsets of the basement/cover interface on these faults appear nearly vertical, suggests that they may have an appreciably higher angle of dip, probably $>45^\circ$. Despite this lack of information on some faults, a possible reason for the excellent definition of the faults in basement could be the presence of fluids in the fault planes generating high acoustic-impedance contrasts. The ODP Shipboard Scientific Party¹² believe that the high localized heat flow in this area is associated with hydrothermal circulation within fault blocks, including basement.

Around the ODP Leg 116 sites and around 81° E 5° S (Fig. 3), where the density of *Darwin* profiles is high, it is possible to follow a few of the faults along strike for as far as ~ 40 km. But, more generally, it is not possible to trace faults between lines only 10 km apart, suggesting a mean fault length of <10 km. Where faults could be followed their strike was found to be 90° E to 100° E. This orientation, which is perpendicular to the strike of fracture zones, the manner in which the faults offset basement and their short length giving an en echelon pattern is reminiscent of the fault fabric found parallel to spreading centres. We thus agree with the interpretation of Weissel *et al.*⁵ that these faults are the result of the reactivation about 7 Myr BP of the pre-existing spreading-centre-formed fault fabric.

From a statistical and structural analysis of the *Darwin* single-channel seismic profiles and comparison with GLORIA data on the fault fabric of a fast-spreading ridge, the East Pacific Rise¹³, Bull¹⁴ demonstrated that the high-angle faults are the result of the reactivation with reverse movement of two sets of pre-existing spreading-centre-formed normal faults. One of these sets of basement faults dips to the north and therefore faces outwards from the original spreading centre, whereas the other set dips to the south and is the more commonly envisaged inward-dipping set. We have imaged the outward-dipping set with the multichannel seismic profiles, dipping at $35\text{--}40^\circ$ throughout the oceanic crust. The inward-facing set, dipping to the south, which is not resolved in the multichannel data, we infer above is likely to have a higher dip in basement.

An important tectonic implication of these observations is the obvious brittle behaviour of the oceanic lithosphere. The resolution of faults penetrating to at least 10-s two-way travel time indicates that the whole crust and possibly the uppermost mantle is deforming under intraplate compression in a brittle manner. Studies of mid-ocean ridge earthquakes^{15–17} have revealed that normal faulting within the median valley occurs on planes that dip at $\sim 45^\circ$. Using the assumption that the centroid depth marks the mean depth of fault slip, it seems that faulting extends from 2 to 10 km below the sea floor for these earthquakes—into the upper mantle. Further evidence for the whole crustal extent of spreading-centre-formed faults is found in the western North Atlantic where White *et al.*¹⁸ image eastward-dipping reflectors which they interpret as representing inward-dipping normal faults. It should be noted, however, that these studies of brittle

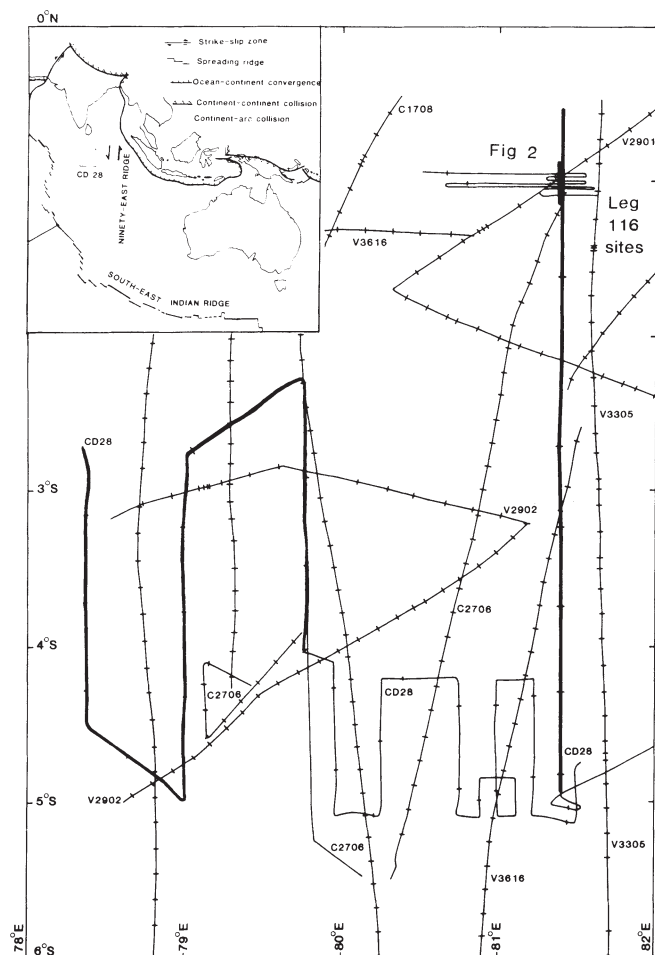


FIG. 1 Position of RRS *Charles Darwin* Cruise 28 in the central Indian Ocean basin together with track chart of seismic profiles which were analysed as part of this work. Plate boundary configuration after Weissel *et al.*⁵ with strike-slip motion along the Ninety-east Ridge as suggested by Stein and Okal²¹. Bold lines in the track chart indicate the position of RRS *Charles Darwin* multichannel lines with the position of Fig. 2 indicated.

failure have been carried out on crust formed at slow spreading rates; little analogous earthquake work has been carried out on fast-spreading ridges, like the East Pacific Rise, similar to the one that produced the central Indian Ocean basin lithosphere. Nevertheless, it is likely that there is a decrease in the thickness of the brittle layer at fast-spreading ridges in view of the evidence from slow-spreading regions that, in general, brittle layer thickness decreases with increasing spreading rate^{15,16}. For large

earthquakes the brittle layer may be as little as 3 km thick when spreading rates of 2.5 cm yr^{-1} are reached¹⁶. Thus it would seem that faulting at the fast-spreading ridge in the Indian Ocean may not have originally penetrated down to Moho and uppermost mantle depths. The deepest parts of the fault planes seen on the multichannel seismic profiles data would then have been formed during reactivation.

Despite the clear evidence for faulting there is a lack of

SHOT POINT NUMBER

900 950 1000 1050 1100 1150 1200 1250 1300 1350

FIG. 2 A short section of fully processed and migrated 12-channel, 12-fold multichannel profile from north (left) to south (right) through the ODP Leg 116 Sites and perpendicular to the structural grain (located A in Fig. 1). Vertical exaggeration is about 3:1 in the sediments and 1.5:1 to 2:1 in basement (the top of which is marked by the strong reflector at ~8-s two-way travel time). Details of the high-angle fault planes in the sediments are evident: some anastomose and show curvature, some are just developing and some are mature. The parallelism of basal sediments and basement suggests basement was originally quite planar and has since been displaced during the intraplate deformation. The fault planes on which the displacement has taken place dip at $30\text{--}40^\circ$ in basement and can be traced to depths at ~10-s two-way travel time (about Moho depth). The multichannel profiles are nearly perpendicular to the strike of the faults and therefore this apparent dip should be close to the true dip. Originally these faults may have been formed at the mid-ocean ridge to the south in which case they must have been outward-dipping faults.

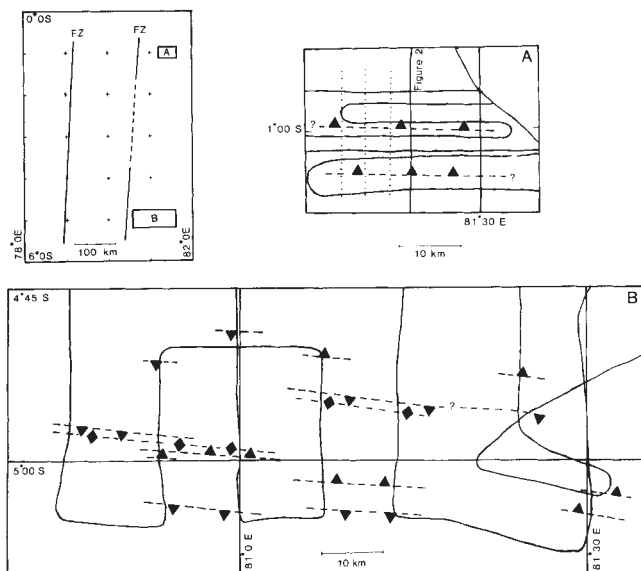
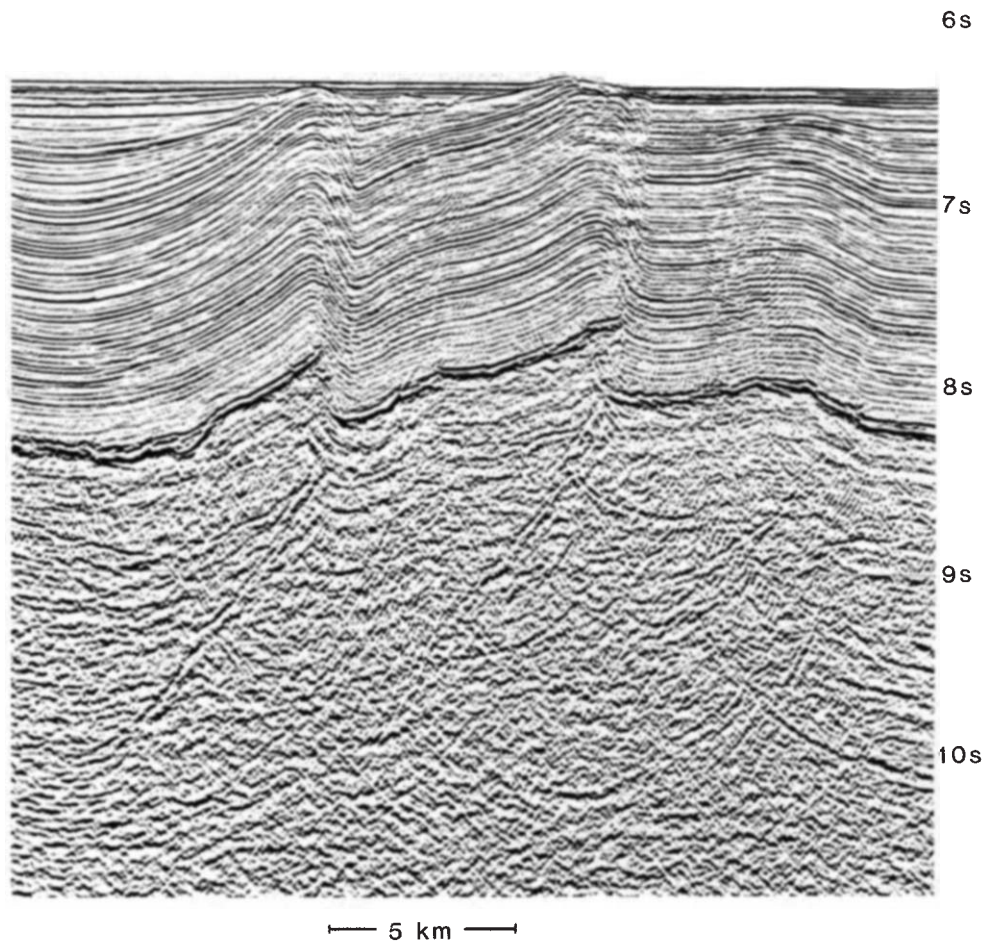


FIG. 3 Plan views of reverse faulting in two areas of the intraplate study area. Top left illustrates the relative positions of the two areas: area A around the ODP Leg 116 sites and area B around $81^\circ\text{E } 5^\circ\text{S}$. The position and orientation of two fracture zones (FZ) developed in the area are also shown. In the detailed maps of area A (top right) and area B (bottom) reverse faults are indicated by dashed lines with a black triangle pointing in the direction of the hanging wall. *Darwin* tracks are continuous lines; in area A the dotted lines are other profiles collected during ODP site surveys and used to constrain the positions of the faults. The position of Fig. 2 is also indicated in area A. In area B the dashed lines with the black diamond ornament show the strike of characteristic hanging-wall anticlines which could be used to tie between lines. It should be noted that, for clarity, not every fault has been annotated in area B: those shown are only the larger ones. In both areas the strike of the reverse faults is 90°E – 100°E , roughly perpendicular to the strike of fracture zones (005°E – 010°E) developed nearby. The general short fault length (usually $<10 \text{ km}$) and resultant en echelon pattern displayed in area B is similar to the fabric developed close to active spreading centres and is consistent with the reverse faults resulting from the reactivation of the pre-existing spreading-centre-formed fabric.

recorded earthquakes associated with the crustal deformation in the central Indian Ocean basin. This could be due to the brittle strength of the oceanic crust and uppermost mantle being too low to sustain significant stresses¹⁹ resulting in only low-magnitude seismicity that is below the level of detection on the worldwide seismograph network. The focal depths of large-magnitude earthquakes recorded in this area are concentrated at depths between 27–39 km (ref. 19), extending down to the expected depth of the brittle/ductile transition. It is interesting that this may be compared with the pattern of continental seismicity in the seismogenic upper crust (see ref. 20, for example) with the largest earthquakes nucleating at or near the base of the brittle layer. By analogy, recent brittle deformation in the central Indian Ocean basin may nucleate at upper-mantle depths and propagate into and through the crust, in the upper part by the reactivation of pre-existing spreading-centre-formed faults. □

Received 28 November 1989; accepted 27 February 1990.

- Bergman, E. A. & Solomon, S. L. *Phys. Earth planet. Int.* **40**, 1–23 (1985).
- Geller, G. A., Weisell, J. K. & Anderson, R. N. *J. geophys. Res.* **88**, 1018–1032 (1983).
- Haxby, W. F. & Weisell, J. K. *J. geophys. Res.* **91**, 3507–3520 (1986).
- Eittrheim, S. L. & Ewing, J. *J. geophys. Res.* **77**, 6413–6421 (1972).
- Weisell, J. K., Anderson, R. N. & Geller, C. A. *Nature* **287**, 284–291 (1980).
- McAdoo, D. C. & Sandwell, D. T. *J. geophys. Res.* **90**, 8563–8569 (1985).
- Zuber, M. T. *J. geophys. Res.* **92**, 4817–4825 (1987).
- Neprochnov, Y. P., Levchenko, O. V., Merklin, L. R. & Sedov, V. V. *Tectonophysics* **156**, 89–106 (1988).
- Cochran, J. R. *et al. Nature* **330**, 519–521 (1987).
- Curry, J. R. & Munasinghe, T. *Earth planet. Sci. Lett.* **94**, 71–77 (1989).
- Leger, G. T. thesis, Dalhousie Univ. (1989).
- Cochran, J. R. *et al. Proc. ODP, Init. Repts* **116**, (College Station, Texas, Ocean Drilling Program, 1989).
- Searle, R. *Tectonophysics* **101**, 319–344 (1984).
- Bull, J. M. *Tectonophysics* (in the press).
- Huang, P. Y. & Solomon, S. C. *J. geophys. Res.* **92**, 1361–1382 (1987).
- Huang, P. Y. & Solomon, S. C. *J. geophys. Res.* **93**, 13445–13477 (1988).
- Toomey, D. R., Solomon, S. C. & Purdy, G. M. *J. geophys. Res.* **93**, 9093–9112 (1988).
- White, R. S. *et al. Geology* (in the press).
- Bergman, E. A. *Tectonophysics* **132**, 1–35 (1986).
- Jackson, J. A. in *Continental Extensional Tectonics* (eds Coward, M. P., Dewey, J. F. & Hancock, P. L.) *Geol. Soc. Spec. Publ.* **28**, 3–17 (1987).
- Stein, S. & Okal, E. A. *J. geophys. Res.* **83**, 2233–2246 (1978).

ACKNOWLEDGEMENTS. We thank the master, officers and crew of RRS *Charles Darwin* and the engineers of RVS *Barry* without whose expertise at sea this paper would not have been possible. We are grateful to J. Cochran and J. Weisell for allowing access to the Lamont-Doherty seismic profiles on which some of our ideas are based, and to R. S. White for critically reading the manuscript. This work was supported by the NERC.

Metabolism of leatherback turtles, gigantothermy, and thermoregulation of dinosaurs

Frank V. Paladino, Michael P. O'Connor* & James R. Spotila*

Department of Biological Sciences, Purdue University, Fort Wayne, Indiana 46805, USA

* Department of Bioscience and Biotechnology, Drexel University, Philadelphia, Pennsylvania 19104, USA

LEATHERBACKS (*Dermochelys coriacea*) are among the largest living reptiles (>900 kg)^{1,2} and range from the tropics to north of the Arctic Circle^{3,4}. They maintain elevated body temperatures (25.5 °C) in cold seawater (7.5 °C)^{5,6} and heat up on land⁷. Metabolic and thermoregulatory mechanisms of leatherbacks have important implications for considerations of size and function in animal biology^{8–10} and for speculation on the endothermic capacities of dinosaurs^{11–18}. Here we report that metabolic rates of adults at rest and while nesting are intermediate to those predicted by allometric relationships for reptiles and mammals. Mathematical modelling indicates that leatherbacks can use large body size, peripheral tissues as insulation, and circulatory changes, to maintain warm temperatures in the North Atlantic and to avoid

overheating in the tropics. This 'gigantothermy' probably allowed large dinosaurs to live in varied habitats, including Cretaceous polar regions.

We measured the resting metabolic rates (RMR) of six adult female leatherbacks (250–430 kg), and their metabolic rates while they were covering their nests or walking, at Tortuguero, Costa Rica in March–April 1989. Metabolism during activity (Fig. 1) ranged from 0.836 to 1.818 W kg⁻¹ ($\bar{x}$ = 1.227 W kg⁻¹). Metabolism in one turtle walking on the beach for 45 min was between (0.836–1.232 W kg⁻¹, $\bar{x}$ = 0.976 W kg⁻¹) the lowest and highest values for nesting animals. Green turtles (*Chelonia mydas*) have slightly higher metabolic rates when covering nests on the beach than when exercising^{19,20}. Resting metabolism ranged from 0.186 to 0.544 W kg⁻¹ ($\bar{x}$ = 0.387 W kg⁻¹). Maximum metabolism, the highest value for each of five nesting leatherbacks, ranged from 0.878 to 1.818 W kg⁻¹ ($\bar{x}$ = 1.510 W kg⁻¹), giving this species a metabolic scope of 4.2, a value similar to that of green turtles^{19,20} and other reptiles.

RMRs of leatherbacks are three times those predicted from allometric relationships for green turtles and other reptiles scaled to leatherback size (Fig. 2)^{19,20}, but are half the values predicted for mammals of this size. Maximum metabolic rates of leatherbacks lie between predicted resting and active metabolic rates for mammals.

We assessed the thermoregulatory capabilities of a leatherback by mathematically analysing heat exchange within the turtle, and between the turtle and the environment (Fig. 3). As body size increases, resting metabolism can maintain progressively larger core-skin temperature differences. With negligible

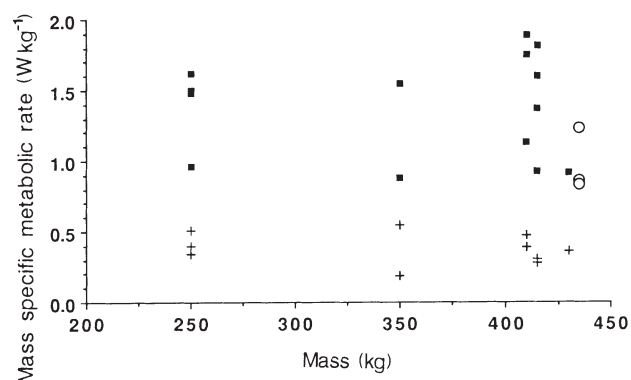


FIG. 1 Mass-specific metabolic rates for nesting leatherbacks at Tortuguero, Costa Rica. Expired respiratory gases were collected and analysed for O₂ and CO₂ concentration, and metabolic rates calculated using data on breathing rate, time and gas volumes. We measured ventilation and gas exchange by analysis of volume and composition of expired air^{19,20}. A 20-l plastic jug with the end removed served as a respiratory mask and was sealed on the turtle's head. A Hans Rudolph giant Y-shaped valve connected to the mask opening separated inspiratory and expiratory flow. The inspiratory side was open to the air and the expiratory side connected by means of a large bore tubing to a 200-l Douglas bag. After collection, the gas was analysed for O₂ using a Beckman C-2 gas analyser, and for CO₂ using a Scholander 0.5-ml gas analyser. We measured volume of expired gas using a dry gas meter. We captured turtles after they nested, restrained them in a cargo net, and weighed them with a large tripod, hoist and scale ($\pm$ 5 kg). After several minutes of restraint, the animals generally became quiescent. Thirty minutes later, we collected expired gases to determine RMR. RMRs may be inflated by whatever aerobic metabolism the turtles used to repay any oxygen debt accumulated during nesting, but other activity was minimal. +, RMRs; ■, metabolic rates of turtles covering their nests; ○, metabolic rates of turtles walking. The lowest value of active metabolism for a turtle was for the period after egg laying when she was covering her nest with her hind flippers. The highest values were when she was vigorously throwing sand with fore and hind flippers, crawling, and rotating the body from side to side while covering the nest with sand. We could not measure gas exchange during egg laying because that would have caused the turtle to abort the nesting process.

heat transfer by means of blood, large animals with mammalian metabolic rates could maintain body cores 30 °C above ambient without fur insulation, whereas small mammals raise metabolism or increase insulation, or both, to maintain similar gradients (Fig. 2). Small reptiles, even while active, will have core temperatures <1 °C above skin temperature. Even a 100-kg reptile, at rest, can maintain only a 3–4 °C core-surface temperature difference. Increased peripheral blood flow would reduce core-skin temperature gradients. Therefore, green turtles (100 kg) have deep body temperatures 1–2 °C above water temperature, except during vigorous swimming, when the pectoral muscles can be 8 °C above water temperature²¹. With minimal blood flow to the skin, leatherbacks could maintain a core-skin temperature difference of 10–20 °C at rest and ≥ 30 °C when swimming.

We used additional simulations to investigate the effect of blood flow on body temperatures. For a leatherback (400 kg), we assume that the temperature in all parts of the body core is uniform owing to blood flow (Fig. 3). As blood flow to the insulating layers increases (Fig. 4a, b), the radius of the core (r_c) approaches the total radius (r_t), and the difference between core (T_c) and environmental (T_e) temperatures decreases. At RMR a leatherback can maintain a 10 °C $T_c - T_e$ temperature difference if it maintains an r_c/r_t ratio of 0.7, that is, if body core is 50% of the total body mass (Fig. 4b). When swimming vigorously (metabolic rate 0.8–1.8 W kg⁻¹), its T_c would be 18–42 °C above water temperature if $r_c/r_t = 0.7$ (Fig. 4a). Even if r_c/r_t is 0.9, $T_c - T_e = 7$ –16 °C. Therefore the turtle should be able to control body temperature in both tropical and temperate seas by changing metabolism and blood flow. A leatherback actively swimming in 5 °C water can maintain a body temperature of 23–29 °C with a metabolism of 0.8 W kg⁻¹ (the lowest active metabolic rates measured in this study). This agrees with data from Frair *et al.*⁶. To avoid overheating while swimming in tropical seas or while nesting, a leatherback would have to perfuse its skin to raise r_c/r_t to near 1.0. That leatherbacks use this mechanism is suggested by our observations that their soft skin turns pink during nesting and appears to be flushed with blood. This was also seen in a leatherback that increased in deep body temperature while held out of water in a cooling environment⁷. We suggest that leatherbacks use changes in blood flow to regulate body temperature and consequently can main-

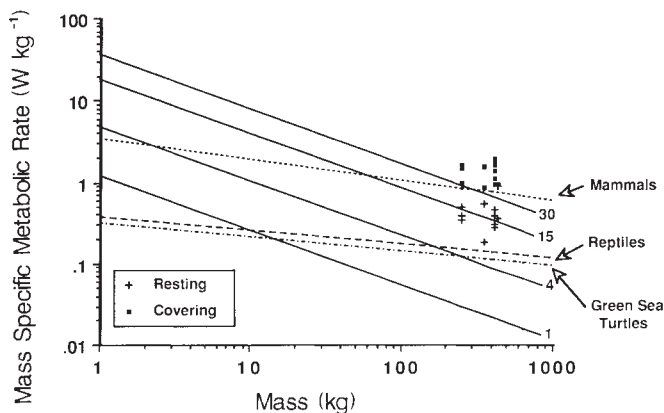


FIG. 2 Comparison of mass-specific metabolic rates for nesting leatherbacks with other species. Leatherback rates were compared with predictions from allometric expressions for resting metabolic rates in lizards (RMR = $0.378 M^{-0.17}$, $T_b = 30$ °C), green turtles (RMR = $0.317 M^{-0.174}$) and mammals (RMR = $3.35 M^{-0.25}$, $T_b = 38$ °C)^{19,20,28}. Also shown (solid lines) are metabolic rates necessary to maintain various differences in temperature (numbers at ends of lines, °C) between the centre and the skin of a cylindrical animal with the thermal characteristics of the turtle. See Fig. 3 for description of model. Blood flow is assumed to play no part in exchanging heat among the different parts of the animal.

tain an elevated body temperature in the cold North Atlantic and avoid overheating in tropical seas and while nesting.

The core-environment temperature differences predicted by our model are larger than those previously predicted²², based on extrapolations of empirical relationships seen in smaller animals. Our simulations and others (ref. 23) indicate that such extrapolation may be overextended.

Small mammals and birds use high metabolic rates and high quality insulation to regulate body temperatures. The large differences in metabolic rates between small mammals and small reptiles, such as lizards, make it appropriate to generalize and label such groups as endothermic homeotherms or ectotherms, respectively. This is not the case for large mammals and large reptiles that have similar body sizes, insulative capacities and

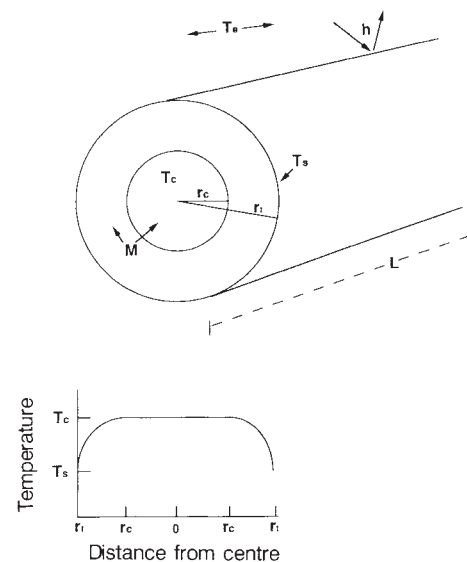


FIG. 3 Structure of the 'core-shell' cylinder model of a leatherback turtle. Body temperatures throughout the core cylinder (radius r_c) are assumed to be homogeneous owing to blood flow. Temperatures in the shell (outside radius r_t) vary owing to conduction of heat between the core and surface. Blood flow to the shell is assumed to be low enough that conduction is the predominant form of heat transfer. Total length of the cylindrical turtle is assumed to be $4r_t$. Heat is exchanged with the thermal environment at the surface of the shell cylinder (heat transfer coefficient h). The difference between body temperatures at the animal's core and those of the environment (either air or water) at thermal equilibrium is then given by:

$$T_c - T_e = \frac{M_t}{A_t h} + \frac{q}{4k} (r_t^2 - r_c^2) - C_1 \ln \left(\frac{r_t}{r_c} \right)$$

Total difference in temperature = Difference due to external resistance + Difference due to conduction through core and shell

where $C_1 = r_c/k(qr_c/2 - M_c/A_c)$, T_c = temperature of body core (°C), T_e = environmental (air or water) temperature (°C), A_c = area of interface between core and shell (m²), A_t = surface area of the shell (m²), M_c = metabolic heat production in the whole animal (W), r_c = radius of central core (m), r_t = radius of the body cylinder (m), q = metabolic heat production per unit volume (W m⁻³), k = thermal conductivity of tissue (W m⁻¹ K⁻¹), h = convection coefficient (W m⁻² K⁻¹). For simulations of turtles on the nesting beach, h is calculated using a wind speed of 2 m s⁻¹. For simulations of turtles in the water, the animal is assumed to be swimming at 3.1 km h⁻¹. The other variables are either characteristics of the animal (for example k) or depend only on the animal's mass and metabolic rate, and the value of r_c/r_t . If blood flow is negligible (that is, no homogeneous core exists), the steady-state temperature difference between core and skin is:

$$\Delta T = \frac{qr^2}{4k}$$

where r = radius of the whole body cylinder (m).

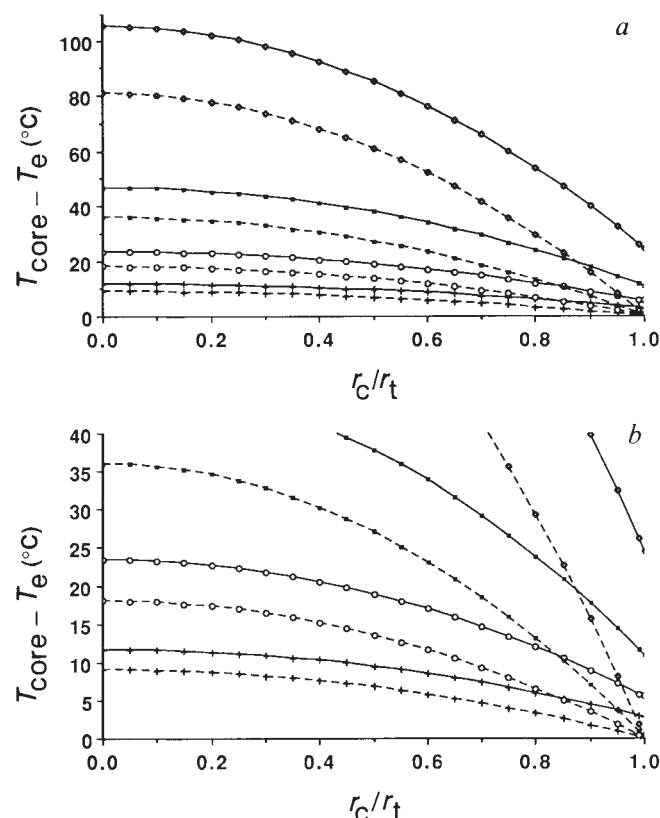


FIG. 4 Differences between core and environmental (air or water) temperatures for a 400 kg turtle at thermal equilibrium with its environment (lower graph shows detail of upper graph). The turtle is assumed to be composed of two concentric cylinders, a central core (with radius r_c) all parts of which are kept at the same temperature by blood flow, and a shell (outside radius $r_t = 32$ cm for a 400-kg turtle) in which blood flow does not affect the flow of heat. Therefore, the proportion of the animal's mass in the core is given by the square of the ratio of the core radius to the total radius (r_c/r_t). Solid lines represent core-air temperature gradients ($T_{\text{core}} - T_{\text{air}}$) on the nesting beach. Dashed lines represent core-water temperature gradients ($T_{\text{core}} - T_{\text{water}}$) for a turtle swimming in water at 3.1 km h^{-1} . Symbols represent different mass-specific metabolic rates: Symbols: +, 0.2 W kg^{-1} ; O, 0.4 W kg^{-1} ; ■, 0.8 W kg^{-1} ; ◇, 1.8 W kg^{-1} .

metabolic rates (Fig. 2). Therefore, we propose the new term 'gigantothermy' to reflect this convergence of thermoregulatory adaptations in large animals and to distinguish this unique thermoregulatory strategy from more traditional explanations of thermoregulation.

Gigantothermy is the maintenance of constant, high body temperatures by means of large body size, low metabolic rate, and use of peripheral tissues as insulation. Large body size obviates thermoregulatory needs for high metabolic rates.

Our experimental data and simulations indicate that gigantothermy would have allowed large dinosaurs with metabolic rates below those predicted for mammals to live in a wide variety of habitats without undue thermal stress, including the polar regions in the late Cretaceous when polar climates were much milder and more equable than they are now^{11,24,25}. Large dinosaurs would not have faced a more severe thermoregulatory challenge than that overcome by leatherbacks in the North Atlantic today. Dinosaurs were probably active, with complex behaviour patterns, including migrations^{11,26,27}. The same is true of living reptiles, and their life styles are readily supported by reptilian physiologies including gigantothermy in leatherback turtles, regional endothermy in green turtles, and behavioural thermoregulation in small reptiles. □

Received 5 September 1989; accepted 27 February 1990.

1. Rhodin, A. G. J., Ogden, J. A. & Conlogue, G. J. *Nature* **290**, 244-246 (1981).
2. Goff, G. P. & Lien, J. *Can. J. Zool.* **102**, 1-5 (1988).
3. Willgoose, J. F. *Nature* **179**, 163-164 (1957).
4. Bleakney, J. S. *Can. J. Zool.* **79**, 120-128 (1965).
5. Mrosovsky, N. & Pritchard, P. C. H. *Copeia* **4**, 624-631 (1971).
6. Frair, W., Ackman, R. G. & Mrosovsky, N. *Science* **177**, 791-793 (1972).
7. Standora, E. A., Spotila, J. R., Keinath, J. A. & Shoop, C. R. *Herpetologica* **40**, 169-176 (1984).
8. Peters, R. H. *The Ecological Implications of Body Size* (Cambridge University Press, Cambridge, 1983).
9. Calder, W. A. *Size, Function and Life History* (Harvard University Press, Cambridge, Massachusetts, 1984).
10. Schmidt-Nielsen, K. *Scaling: Why is Animal Size So Important* (Cambridge University Press, New York, 1984).
11. Paul, G. S. *J. Paleontol.* **62**, 640-652 (1988).
12. Spotila, J. R., Lommen, P. W., Bakken, G. S. & Gates, D. M. *Am. Nat.* **107**, 391-404 (1973).
13. Spotila, J. R. in *A Cold Look at the Warm-Blooded Dinosaurs* (eds Thomas, R. D. K. & Olson, E. C.) 233-254 (AAAS Symposium 28, Westview Press, Boulder, Colorado, 1980).
14. Feduccia, A. *Evolution* **27**, 166-169 (1973).
15. Bennett, A. F. & Dalzell, B. *Evolution* **27**, 170-174 (1973).
16. Dodson, P. *Evolution* **28**, 494-497 (1974).
17. Farlow, J. O. *Ecology* **57**, 841-857 (1976).
18. McGowan, C. *Paleobiology* **5**, 285-295 (1979).
19. Prange, H. D. & Jackson, D. C. *Resp. Physiol.* **27**, 369-377 (1976).
20. Jackson, D. C. & Prange, H. D. *J. comp. Physiol.* **134**, 315-319 (1979).
21. Standora, E. A., Spotila, J. R. & Foley, R. E. *J. thermal Biol.* **7**, 159-165 (1982).
22. Alexander, R. M. *Dynamics of Dinosaurs and Other Extinct Giants* (Columbia University Press, New York, 1989).
23. McNab, B. K. & Auffenberg, W. *Comp. Biochem. Physiol.* **A55**, 345-350 (1976).
24. Brouwers, E. M. et al. *Science* **237**, 1608-1610 (1987).
25. Rich, P. V. et al. *Science* **242**, 1403-1406 (1988).
26. Thomas, R. D. K. & Olson, E. C. (eds) *A Cold Look at the Warm-Blooded Dinosaurs* (AAAS Symposium 28, Westview Press, Boulder, Colorado, 1980).
27. Farlow, J. O. & Dodson, P. *Evolution* **29**, 353-361 (1975).
28. Bennett, A. F. in *Biology of the Reptilia* Vol. 13 (eds Gans, C. & Pough, F. H.) 155-199 (Academic, New York, 1982).

ACKNOWLEDGEMENTS. We thank the following for technical assistance: J. P. Ross, R. E. Gatten, Jr, G. Cruz, A. Marchfelder, E. Gutierrez, C. Roldan Chacon, A. Chaves, B. Bowen, J. Avise, J. A. Spotila, D. N. Penick, M. and P. Faucher, D. Hill, D. Peck, R. and E. Worth, B. T. and H. C. Greenwalt, L. and P. George, D. Brooks, S. Austin, B. A. Peters, G. Wagner, B. Mentzer, and J. and B. Fensterwald. Protocols were approved by animal care committees of Drexel and Purdue Universities. This work was supported by the Caribbean Conservation Corporation and the DOE.

Population estimates for the dunlin *Calidris alpina* derived from remotely sensed satellite imagery of the Flow Country of northern Scotland

M. I. Avery & R. H. Haines-Young*

Research Department, Royal Society for the Protection of Birds, The Lodge, Sandy, Bedfordshire SG19 2DL, UK

THE Flow Country of northern Scotland is of international importance for wading birds, and has in the last decade been greatly affected by afforestation^{1,2}. Here we demonstrate that remotely sensed satellite imagery of this area can be used to make accurate predictions of the numbers of the dunlin *Calidris alpina*, and thereby assess the impact of forestry on this wading bird. The habitat favoured by dunlin is wet moorland interspersed with small pools; they do not use large water bodies, built-up areas, woodland or agricultural land. We show first that the numbers of dunlin on 42 moorland sites, as surveyed in 1986 (ref. 3), are significantly negatively correlated to an index of soil wetness⁴, as provided by reflectance data from a satellite image of these sites taken in 1978 before afforestation. We then tested this relationship by using it to predict dunlin numbers on a random sample of unsurveyed sites. Subsequent surveying of these sites in 1988 revealed a good fit between predicted and observed numbers, and confirms previous independent estimates of dunlin loss⁴.

The near-infrared band 7 of Landsat MSS is sensitive to variations in vegetation and ground wetness⁴. By using a mask

* Institute of Engineering Surveying and Space Geodesy, University of Nottingham, Nottingham NG7 2RD, UK.

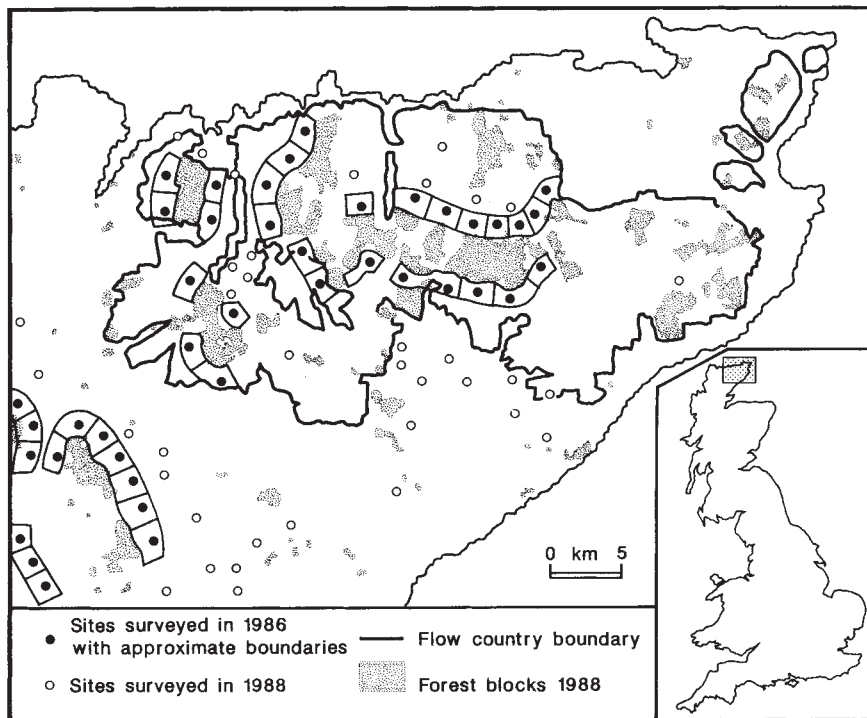


FIG. 1 Map of Flow Country indicating sites of dunlin counts in 1986 and 1988. Approximate boundaries of sites surveyed in 1986 are shown. Locations of 2.5 km x 2.5 km test sites surveyed in 1988 are indicated by their centroid.

to remove all the areas in the Flow Country which would be unsuitable for dunlin, we hypothesized that remotely sensed data would provide an index of habitat quality. Because wet sites produce low band-7 reflectance and dry sites produce higher values⁴, we also hypothesized that there should be an inverse relationship between the mean infrared reflectance for the site and the numbers of dunlin.

We tested these hypotheses with data on dunlin numbers collected in 1986 from transect surveys of 42 moorland sites³, each 4.0 km x 1.6 km (Fig. 1). A linear regression model was constructed, relating population numbers at a site to mean band-7 reflectance derived from the satellite image of 8 May 1978. This image was used because it was cloud-free and pre-dated most afforestation. The null hypothesis of no relationship

was rejected at the 95% significance level (Fig. 2).

We used this fitted regression to predict dunlin numbers for each 0.25 km x 0.25 km cell in the study area that had not been assigned to one of the unsuitable cover types by the mask. The cells were combined to produce predictions on a grid with cells of 2.5 km x 2.5 km. A random set, stratified by expected dunlin numbers, of 37 such squares was surveyed in 1988 to test the predictions of the model. The sample was restricted to cells of >90% moorland cover to make the test more stringent by excluding inherently poor areas from the survey. The field team was not aware of the predictions made for each site. The transect lines were at 500-m intervals so that, because only birds within 100-m of the line were recorded (as in 1986), 40% of the area of each site was covered. In this way, we could maximize the

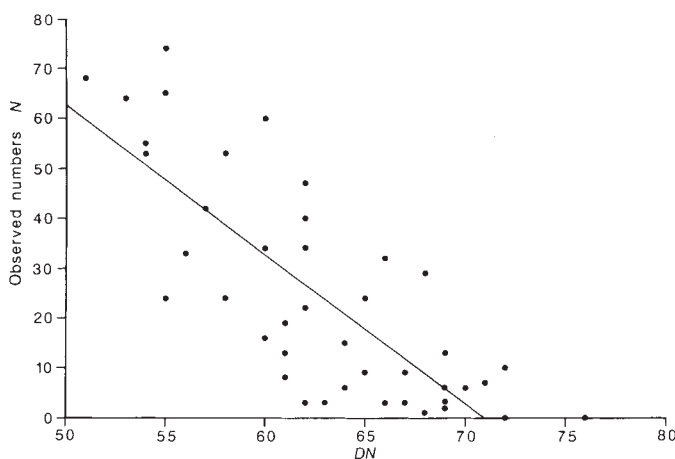


FIG. 2 Observed dunlin numbers (N) plotted against mean MSS satellite band-7 digital number (DN). $N = 209.7 - 2.94 \times DN$; coefficient of correlation $r = -0.79$ (significant at the $P = 0.05$ level). Dunlin numbers are the sum of all birds seen in 1986 in two visits to each site, one in May, and the other in June.

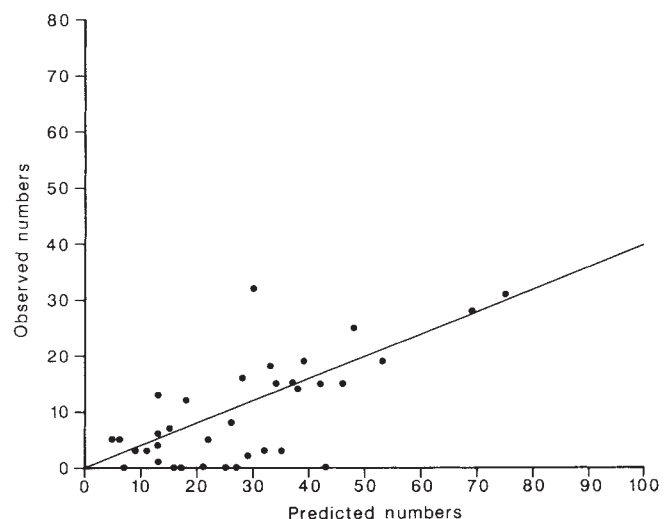


FIG. 3 Predicted dunlin numbers plotted against observed numbers for sites in 1988. Sites were surveyed at 40% of the intensity of the 1986 survey; therefore the predicted slope is 0.4 (plotted for reference), with an intercept at (0, 0). Observed slope of regression was $+0.39 \pm 0.07$; observed intercept, -1.7 ± 2.12 .

number of sites able to be surveyed with the team available. There was no significant difference between the observed and predicted numbers (Fig. 3).

Our method allows the effects of habitat loss on dunlin to be assessed. Between 1978 and 1988, 17% of the Flow Country was lost to afforestation. From the infrared reflectances of the areas afforested, we conclude that dunlin populations have declined by 18%. This compares with independent estimates² of a 19% decline for the same general period.

Because the relationship established between an image from 1978 and dunlin numbers in 1986 remains valid in 1988, the application of similar techniques to similar problems may not depend on the availability of recent imagery. □

Received 4 September 1989; accepted 19 February 1990.

1. Stroud, D. A., Reed, T. M., Pienkowski, M. W. & Lindsay, R. A. *Birds, Bogs and Forestry* (Nature Conservancy Council, Peterborough, 1987).
2. Bainbridge, I. P., Housden, S. D., Minns, D. & Lance, A. N. *Forestry in the Flows of Caithness and Sutherland. Conservation Topic Paper No. 18* (Royal Society for the Protection of Birds, Sandy, UK, 1987).
3. Avery, M. I. *J. appl. Ecol.* **26**, 957–966 (1989).
4. Mather, P. M. *Computer Processing of Remotely Sensed Images* (Wiley, London, 1988).

ACKNOWLEDGEMENTS. This work was funded in part by the Forestry Commission. We thank Colin Bibby for inspiration and Frances Winder for helping to prepare the mask.

High productivity of northern Bering Sea benthic amphipods

Raymond C. Highsmith & Kenneth O. Coyle

Institute of Marine Science, University of Alaska Fairbanks, Fairbanks, Alaska 99775-1080, USA

POLAR regions, particularly Arctic seas covered by ice for a significant portion of the year, have been considered to be relatively unproductive^{1–4}. In the western Arctic, including the northern Bering Sea, high water column productivity^{5,6}, and well developed benthic communities^{7–9} have recently been reported. Ampeliscid amphipod crustaceans are the benthic community dominants in vast areas of the northern Bering Sea and are the major prey of the migratory California grey whale, *Eschrichtius robustus*^{10,11}. Our study indicates that productivity of the benthic amphipods is remarkably high, given the latitude. The dominant species, *Ampelisca macrocephala*, is the most productive benthic marine amphipod ever reported and the amphipod community is more productive than entire benthic communities investigated elsewhere. Polar marine ecosystems clearly have the potential to be highly productive.

Here we develop estimates of productivity of the northern Bering Sea benthic community, and make comparisons with other benthic communities around the world, in preparation for assessing the importance of ampeliscid amphipods and other invertebrates to the energetics of grey whales.

Our study area, bounded by 64° to 65° N and 168° to 170° W, is located in the Chirikov Basin between St. Lawrence Island and Bering Strait (Fig. 1). Estimates of the area occupied by the extensive amphipod community range from 2.2×10^4 (ref. 12, 13) to 3.75×10^4 km² (ref. 11). The region is typically ice-covered for five to six months each year during the oceanographic winter. This is presumably a period of little or no primary production and, hence, probable food shortage for the diatom-eating amphipods. In spite of very cold sea-floor water temperatures ranging annually between -1.7 to $+2.5$ °C⁹ (unpublished observations) and the limited primary production season, we hypothesized that the amphipod community must be extraordinarily productive in order to support intensive predation by the grey whales.

The Chirikov Basin is a region of seasonally high primary production with annual rates of about 250 – 325 g C m⁻² (refs 5, 6, 14). As water depths are only 35 to 50 m, a substantial

portion of the water column production is apparently ungrazed by zooplankton and settles to the bottom⁸, thus becoming available to the benthic, tube-building amphipods. Phytoplankton counts in our water samples collected 0.5–1 m above the bottom indicate that algal cells, particularly diatoms, were settling out in high concentrations of up to 37 million cells per litre ($N = 120$ samples). The amphipods position themselves ventral side

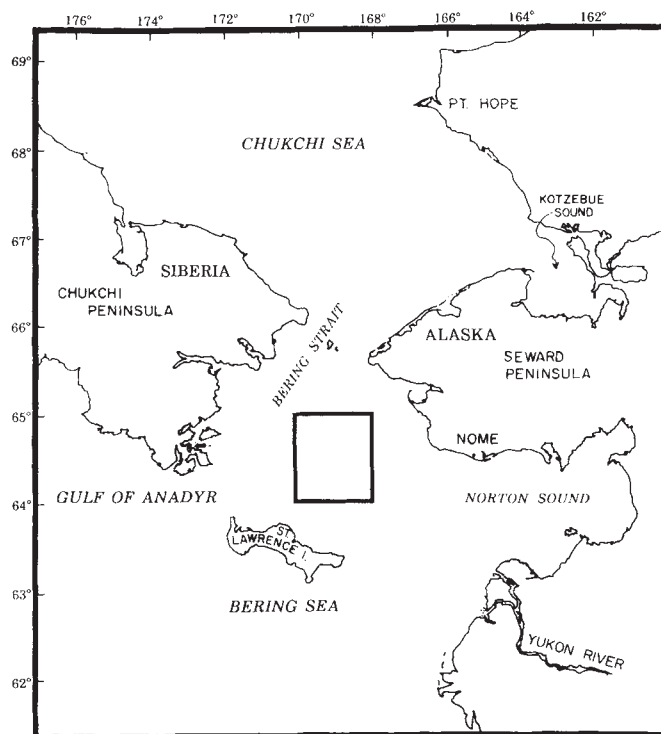


FIG. 1 Location of the Chirikov Basin study area (rectangle) in the northern Bering Sea.

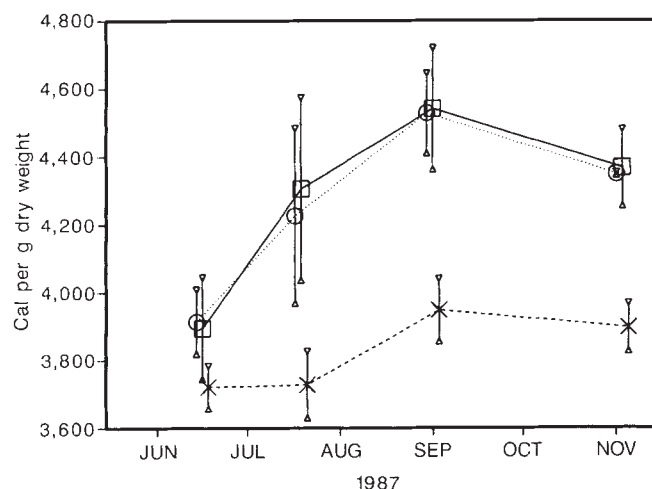


FIG. 2 Mean caloric values for *A. macrocephala* (□), *A. birulai* (×) and *Byblis* spp. (○) during the ice-free period, 1987. Vertical bars indicate standard deviation. Within species, there was no significant correlation between caloric content and animal length. Total number of bombings of homogenates of several animals was 136, 20 and 51, respectively. Measurements were made with a Parr Bomb Calorimeter with semi-micro attachment, using standard procedures.

TABLE 1 Mean amphipod abundance and dry weight at 16 stations in the Chirikov Basin

Species	Number per m ²	g m ⁻²
<i>Ampelisca macrocephala</i>	2,499 (±1,183)	42.6 (±15.5)
<i>A. birulai</i>	1,068 (±646)	3.0 (±1.9)
<i>Byblis</i> spp.*	1,039 (±912)	7.1 (±4.3)
Totals	4,606	52.7

Standard deviation in parentheses. Five van Veen grab samples were collected per station in May, July and September 1986 ($N=240$ samples). Samples were rinsed through a 1-mm mesh screen.

* It is not possible to identify specimens of *Byblis* less than about 7 mm in length to species, so pooled data are presented for *B. gaimardi*, *pearcyi*, *frigidis*, *robustus* and *breviramus*. *B. gaimardi* is much more abundant than the other species, followed by *frigidis* and *pearcyi*. *B. frigidis* and *B. robustus* are new species²⁴.

up at the entrance to their tubes and generate a posterior to anterior feeding current (revealed by food colouring added to the water) and, thus can feed on the diatom suspensions. In addition, we observed that deposited materials on the surrounding sediment surface can be swept into the feeding current by the antennae.

A. macrocephala, a relatively large amphipod species with a maximum length of about 3 cm, makes up 81% of the total ampeliscid biomass in the study area (Table 1). *A. birulai* is the second most abundant species but, because of its relatively small size (12–14 mm), accounts for only about 6% of the biomass. The remaining ampeliscid biomass reported here is composed of a group of five species of *Byblis*. Collectively, these ampeliscids account for nearly all of the infaunal amphipod biomass in our samples.

The caloric content per unit weight of the amphipods increased during the summer (Fig. 2), suggesting accumulation of energy-rich storage compounds such as lipids in preparation for winter.

Standing crop in kcal m⁻² (Table 2a) was calculated from biomass (Table 1) and caloric data (Fig. 2). Production rates (Table 2a) were estimated from length-frequency data using the cohort summation technique^{15,16}. The amphipod production period is presumably related to the ice-free period of planktonic primary production, around May to November. Based on the relatively high caloric values from June to November (Fig. 2), the annual period of amphipod production appears to be at least six months long and cohort analysis of growth rates suggests an eight-month period. The following mechanisms could extend the amphipod production period beyond the approximate six-month primary production season: feeding on diatoms advected northward from the spring bloom associated with break-up when the ice edge is south of the study region (there is a net northward flow of water through Bering Strait¹⁷); feeding during the winter on settled diatoms accumulated in surface sediments; and/or feeding on alternative food sources such as detritus, zooplankton and larvae¹⁸. An amphipod production period of six to eight months yields production/dry weight biomass ratios between 0.8 and 1.0 (Table 2a). The low P/B ratio (turnover rate) indicates that the high ampeliscid production values in the Chirikov Basin result from high standing crop rather than high growth or reproduction rates. Production/biomass ratios for temperate amphipod communities typically fall between 2 and 5 (refs 16, 19–22).

A. macrocephala is the most productive benthic marine amphipod species yet reported (Table 2b). The next most productive sublittoral species is *A. armoricana* from the Bay of Morlaix in the western English Channel²³, with a dry weight production of 24–25 g m⁻² (about 100 kcal m⁻²). Combined ampeliscid production rates reported here for the northern Bering Sea are higher than total benthic community (including meiofauna)

TABLE 2 Benthic biomass and production

a, Amphipod biomass and production estimates for the major ampeliscid species in the Chirikov Basin

Species	Biomass (kcal m ⁻²)	Production (kcal m ⁻²)*	
		6 months	8 months
<i>Ampelisca macrocephala</i>	181	127	169
<i>A. birulai</i>	11	14	19
<i>Byblis</i> spp.	30	31	41
Totals	222	172	229
P/B		0.8	1.0

b, Comparison of most productive amphipod species reported

Species	Location	Production (g dry wt m ⁻² yr ⁻¹)	Source
<i>A. macrocephala</i>	Bering Sea	30–40	This study
<i>A. armoricana</i>	English Channel	24–25†	19
<i>Eogammarus confervicolus</i>	Squamish Estuary	21–22	18
<i>A. araucana</i>	Chile	8–10	17

c, Comparison of Chirikov Basin amphipod production with total benthic production at other locations

Location	Production (kcal m ⁻²)	Source
Chirikov Basin	170–230	This study
Long Island Sound	120	25
Georges Bank	110	26
North Sea	70	27
Scotian Shelf	54	27
Baltic Sea	27	28

* Caloric values based on dry weights (Table 1) and measurements of caloric content per unit weight (Fig. 2).

† Highest previously reported for sublittoral species. *E. confervicolus* is an intertidal, estuarine species.

production estimates for Long Island Sound or Georges Bank, some of the highest published to date (Table 2c). Clearly, the amphipod community, inhabiting tens of thousands of square kilometres of the Chirikov Basin, is one of the most productive benthic communities in the world, even though the region is covered with ice for several months each year. □

Received 11 November 1989; accepted 20 February 1990.

- Conover, R. in *Zoogeography and Diversity of Plankton* (ed. Van der Spoel, S. & Pierrot-Butts, A.) 50–86 (Halsted, New York, 1979).
- George, R. in *Polar Oceans* (ed. Dunbar, M.) 391–408 (Arctic Institutes of North America, Calgary, 1977).
- Knox, G. & Lowry, J. in *Polar Oceans* (ed. Dunbar, M.) 423–462 (Arctic Institute of North America, Calgary, 1977).
- Raymont, J. *Plankton and Productivity in the Oceans* Vol. 1 2nd edn (Pergamon, London, 1980).
- Sambrotto, R., Goering, J. & McRoy, C. *Science* **225**, 1147–1150 (1984).
- Walsh, J. et al. in *Proceedings of the Sixth Conference of the Comité Arctique International*, 13–15 May 1985 (eds Rey, L. & Alexander, V.) 90–120 (Brill, Leiden, 1989).
- Stoker, S. in *The Eastern Bering Sea Shelf: Oceanography and Resources*, Vol. 2 (eds Hood, D. W. & Calder, J. A.) 1069–1090 (University of Washington Press, Seattle, 1981).
- Grebmeier, J., McRoy, C. & Feder, H. *Mar. Ecol. Prog. Ser.* **48**, 57–67 (1988).
- Grebmeier, J., Feder, H. & McRoy, C. *Mar. Ecol. Prog. Ser.* **51**, 253–268 (1989).
- Bogoslovskaya, L., Votrogov, L. & Semenova, T. *Rep. Int. Whal. Commn* **31**, 507–510 (1981).
- Nerini, M. in *The Gray Whale Eschrichtius robustus* (eds Jones, M. L., Swartz, S. L. & Leatherwood, S.) 423–450 (Academic, New York, 1984).
- Johnson, K. & Nelson, C. *Science* **225**, 1150–1152 (1984).
- Nelson, C., Johnson, K. & Barber, J. Jr *J. Sed. Petrol.* **57**, 419–430 (1987).
- Springer, A. Ph.D. dissertation, Univ. Alaska, Fairbanks (1988).
- Crisp, D. in *Methods for the Study of Marine Benthos* (eds Holme, N. & McIntyre, A.) 284–372 (Blackwell, Oxford, 1984).
- Wildish, D. & Peer, D. *Can. J. Fish. Aquat. Sci.* **38**, 1019–1025 (1981).
- Coachman, L., Aagaard, K. & Tripp, R. *Bering Strait: The Regional Physical Oceanography* (University of Washington Press, Seattle, 1975).
- Kanneworff, E. *Ophelia* **2**, 305–318 (1965).
- Collie, J. *Mar. Ecol. Prog. Ser.* **22**, 229–238 (1985).
- Carrasco, F. & Arcos, D. *Mar. Ecol. Prog. Ser.* **14**, 245–252 (1984).
- Stanhope, M. & Levings, C. *Can. J. Fish. Aquat. Sci.* **42**, 1733–1740 (1985).
- Dauvin, J.-C. *J. exp. mar. Biol. Ecol.* **128**, 31–56 (1989).
- Dauvin, J.-C. *J. exp. mar. Biol. Ecol.* **123**, 235–252 (1988).
- Coyle, K. & Highsmith, R. *J. crust. Biol.* **9**, 157–175 (1989).
- Sanders, H. *Bull. Bingham Oceanogr. Coll.* **15**, 344–414 (1956).
- Steimle, F. in *Georges Bank* (eds Backus, R. H. & Bourne, D. W.) 310–314 (MIT Press, Cambridge, MA, 1987).

27. Parsons, T., Takahashi, M. & Hargrave, B. *Biological Oceanographic Processes* (Pergamon, Oxford, 1984).
28. Mann, K. *Ecology of Coastal Waters: A Systems Approach* (University of California Press, Berkeley, 1982).

ACKNOWLEDGEMENTS. We thank for their help Carol Winiecke, Tama Rucker, J. McDonald and A. Paul; several students, including F. Adams, A. Blanchard, J. Carpenter, S. Clay, M. Derenoff, D. Ferris, S. Fox, S. Hartz, T. Henry, H. Kim, B. Leal, G. Owens, P. Ribbens, F. Rivers, C. Saccheus, L. Shelley, C. Sullivan, S. Viereck, J. Warburton, and K. Weller; the entire crew of the R/V *Alpha Helix*; and marine technicians D. Boisseau, S. Ellis, D. Leech and M. Testarmata. This research was supported by National Science Foundation, Division of Polar Programs.

Red Queen hypothesis supported by parasitism in sexual and clonal fish

Curtis M. Lively*, Clark Craddock & Robert C. Vrijenhoek

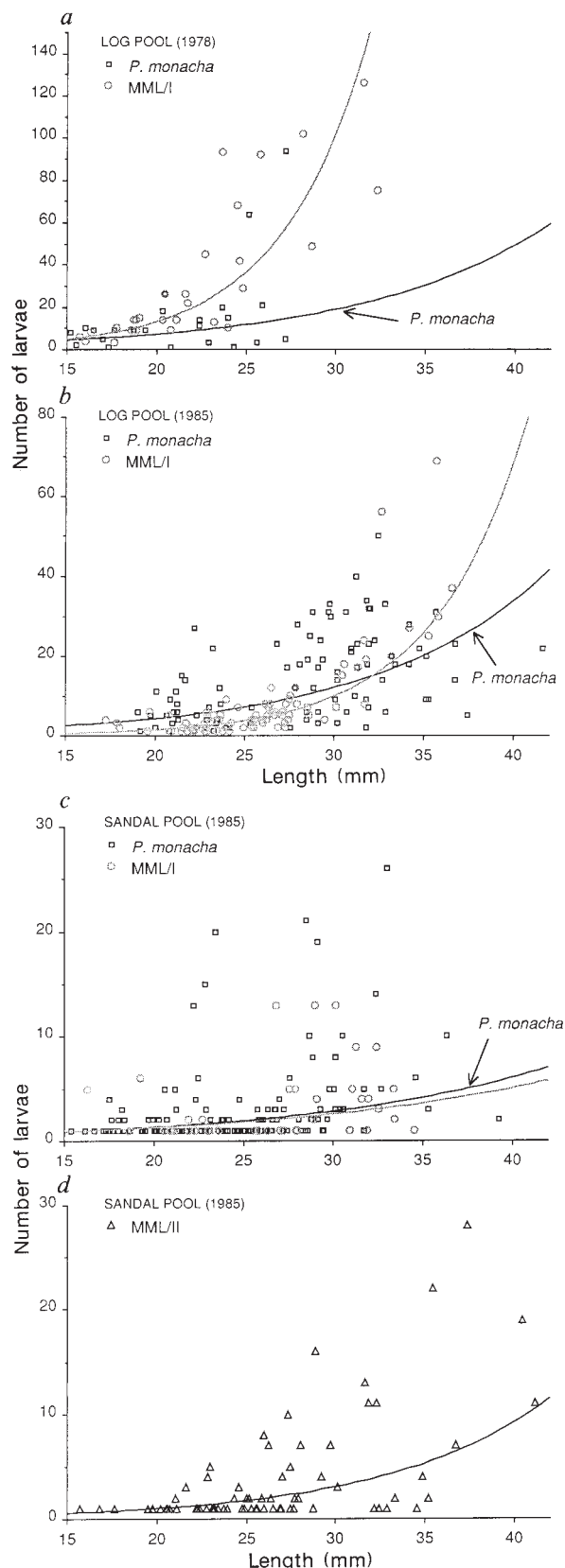
Center for Theoretical and Applied Genetics, Cook College, Rutgers University, New Brunswick, New Jersey 08903-0231, USA

THE Red Queen hypothesis for the maintenance of biparental sexual reproduction suggests that, for species locked in coevolutionary struggles with biological enemies, the production of variable progeny compensates for the genetic or ecological disadvantages of sex¹⁻⁷. The advantage of sex and recombination under this hypothesis stems from the production of rare phenotypes, which are expected to be more likely to escape infection or predation by coevolved biological enemies. Like many evolutionary hypotheses, the Red Queen hypothesis is difficult to test directly, but its assumptions and predictions can be evaluated⁷⁻¹⁸. The most critical assumption is that biological enemies will disproportionately attack the most common phenotype^{19,20}. In this study of parasite loads of coexisting sexual and clonal fish, we find empirical support for this assumption.

The sexual topminnow *Poeciliopsis monacha* (Atheriniformes: Poeciliidae) coexists with two gynogenetic triploid clones of *P. monacha-lucida*, which were formed by hybridization between *P. monacha* and *P. lucida*²¹. The clones are referred to as MML/I and MML/II, and they can be distinguished from each other and from sexual individuals by gel electrophoresis of their proteins and by histocompatibility criteria^{22,23}. These fish commonly exhibit black spot disease, an infection by trematode larvae (*Uvulifer* sp.) which form externally visible cysts after burrowing into the body wall²⁴. In what follows, we assume that susceptibility to helminth infections of this type is genetically based (reviewed in ref. 25) and that the most common susceptibility phenotype in an outbred sexual population is far less frequent than the most common clone. The Red Queen hypothesis therefore predicts that the more common clone will be more intensely infected than coexisting sexual populations.

We counted the number of cysts infecting fish from three natural rock pools in the Arroyo de las Platanos (Río Fuerte, Sonora, Mexico). The first of these (Log Pool) contains *P. monacha* and MML/I. In the two years in which collections were available for this pool (1978 and 1985), clone MML/I accumulated cysts at a significantly greater rate per unit body length than sexual fish (Fig. 1a and b; Table 1).

A similar result was obtained in the second pool (Sandal Pool, located 274 m below Log Pool), in which sexual *P. monacha* coexist with both MML/I and MML/II. The more common clone in this pool, MML/II, was significantly more parasitized per unit body length than either *P. monacha* or MML/I (Fig. 1c and d; legend to Table 1). We observed no significant difference in the infection levels between the less frequent clone MML/I and *P. monacha* (Table 1), suggesting that the previous result for Log Pool was not an artefact of MML/I being



inherently more susceptible than *P. monacha*, independent of its relative abundance.

In the third pool (Heart Pool, 150 m above Log Pool), a local population extinction and subsequent founder event allowed us to compare parasite loads in MML/I with both homozygous and genetically variable *P. monacha*. Heart Pool dried completely during a severe drought in 1976. The fish recolonized

* Present address: Department of Biology, Indiana University, Bloomington, Indiana 47405, USA.

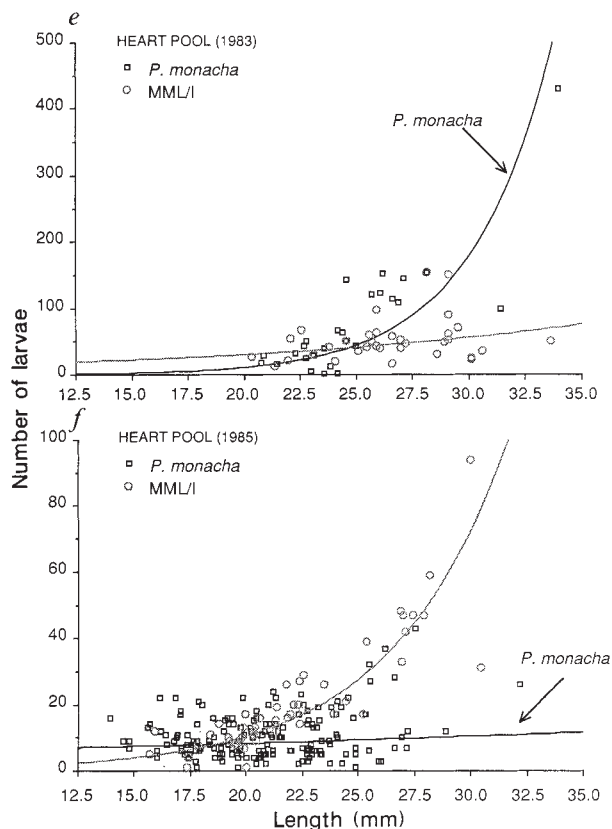


FIG. 1 Bivariate plots of number of encysted trematode larvae (metacercariae) in topminnows against standard length (mm) of the fish. Plots a and b are for Log Pool for 1978 and 1985. Plots c and d are for Sandal Pool in 1985: c gives the results for MML/I and *P. monacha* in the same form as in the previous plots; d gives the results for MML/II. Plots e and f are for Heart Pool: e shows the results for 1983, when the sexual population was highly inbred (removing the 'outlier' of over 400 parasites in this plot does not alter the conclusion that the two slopes are significantly different); f shows the results for 1985, two years after the introduction of sexual females from the Jaguarí river. The formal statistical analysis is given in Table 1. The curves approximate the best fit line from the back-transformed log-linear plots.

the pool by 1978, but genotypic diversity in the inbred sexual founders was severely reduced (changes in gene diversity and population dynamics of these populations are discussed in refs 26 and 27). In 1983, genetically variable sexual females were transplanted from a nearby downstream population, resulting in an increase in the local genotypic diversity of the sexual population²⁶. Collections of the fish from Heart Pool were made both before (1983) and after (1985) the transplant.

Before the transfer of sexual females into Heart Pool, the number of parasites per unit length of fish was significantly greater in the highly homozygous inbred sexual fish than in clone MML/I (Fig. 1e; Table 1). Because MML/I was more common than sexual *P. monacha*, this result is inconsistent with the prediction of greater infection of the most common phenotype. But the higher infection of the inbred sexual individuals might have been due to the detrimental effects of inbreeding depression; inbreeding in this pool resulted in a decline of developmental stability and other fitness-related traits^{26,27}. Two years after the transfer of sexual females to Heart Pool, the trend had reversed and the number of parasites per unit length of fish was significantly greater in clonal fish than in the outbred sexual fish (Fig. 1f; Table 1).

If overall levels of genotypic variability are correlated with variance in disease susceptibility, outbred sexual individuals should exhibit greater variance in parasite loads than both the

TABLE 1 Analysis of covariance summary

Contrast	Pool	Year	Mean sum of squares Interaction (d.f.)	Error (d.f.)	F	P
1	Log	1978	1.21 (1)	0.201 (48)	6.02	0.018
2		1985	1.14 (1)	0.114 (171)	10.06	0.002
3	Heart	1983	3.81 (1)	0.291 (57)	13.08	<0.001
4		1985	5.31 (1)	0.094 (215)	56.60	<0.001
5	Sandal	1985	0.17 (2)	0.054 (251)	3.20	0.042

Analysis of covariance summary table for the heterogeneity of slopes model²⁹. The dependent variable, number of trematode cysts in the body wall, was log-transformed³⁰ before the analysis: $n' = \log(n + 10)$. The independent variables were reproductive mode (sexual versus asexual) and standard length of the fish (covariate). A significant F value in the table indicates an interaction between reproductive mode and standard length, meaning the slopes of the regression lines for parasite number against length are significantly different. We interpret significant differences to reflect differences in the rates of accumulation of parasites as a function of age of the fish. We feel that larger fish show more of a difference between strains because they have had more time to accumulate their parasite loads. (Because both clones are pseudogamous, we do not intend to imply from this analysis that the trematode is selecting for the maintenance of sex in these fish.) The comparisons were between *P. monacha* and MML/I in all cases except Sandal Pool, wherein *P. monacha*, MML/I, and MML/II were compared. For Sandal Pool, the more common clone, MML/II, had a significantly greater slope than *P. monacha* and MML/I ($F_{1,253} = 6.57$; $P = 0.011$); the difference between *P. monacha* and MML/I was not significant ($F_{1,178} = 0.21$; $P = 0.648$). Because the assumption of homoscedasticity was not met in contrasts 1, 2 and 4 (see Table 2), the probability values given here are only approximate.

TABLE 2 Variance ratios

Pool	Year	Residual variance Numerator (d.f.)	Denominator (d.f.)	F ratio	P
Log	1978	0.051 (23)	0.025 (28)	2.04	<0.050
	1985	0.029 (90)	0.012 (80)	2.42	<0.001
Heart	1983	0.078 (26)	0.084 (33)	0.93	0.571
	1985	0.020 (170)	0.010 (47)	2.00	<0.005
Sandal	1985				
	<i>P. monacha</i> , MML/I	0.010 (137)	0.010 (43)	1.00	0.483
	<i>P. monacha</i> , MML/II	0.010 (137)	0.011 (73)	0.91	0.307
	MML/I, MML/II	0.010 (73)	0.011 (43)	0.91	0.341

The residual variance for the sexual population divided by the residual variance for MML/I (F ratio), except in Sandal Pool as indicated. The residuals were calculated from the regression of the number of cysts (transformed to $\log(\text{cysts} + 10)$) against standard length of the fish. Significant differences imply a greater genetic variation for susceptibility to infection in the sexual subpopulation. Such differences, however, are also in violation of the homogeneity of variance assumption for the ANCOVA in Table 1, to which the analysis is normally considered to be robust²⁹.

inbred founder individuals and the genetically uniform clones. We examined this hypothesis by comparing the residual variances in parasite loads (Table 2). For both of the Log Pool samples and the 1985 Heart Pool sample, genetically variable sexual fish exhibited twice the variance in infection levels of the coexisting clone (a highly significant ratio, see Table 2). The homozygous founder population of *P. monacha* from Heart Pool in 1983 did not differ significantly from the clone in this regard. Hence, the greater variation in parasite loads in sexual fish seems to be associated with outbreeding. The parasite load of genotypically variable sexual individuals from Sandal Pool was not significantly more variable than those of the coexisting clonal forms. The reasons for this discrepancy with the other pools are unknown but might be due to the low levels of infection in Sandal Pool (Fig. 1).

In summary, the *Poeciliopsis* system provided an opportunity to examine the assumptions of the Red Queen hypothesis in a natural ecological setting, because it represents a rare situation where clonal and non-clonal forms are recognizable and coexist. We found that the more common clone accumulated parasites at a significantly greater rate than the sexual subpopulation, except in the case where the sexual subpopulation was highly inbred following a founder event. This result corroborates agriculture's experience with monocultures²⁸ and supports an important assumption of the Red Queen hypothesis: parasites

evolve to disproportionately infect genetically uniform strains as they become common. The results are also consistent with our assumption of a genetic basis to the host-parasite interaction, and with the expectation that sexual reproduction provides a partial escape from biological enemies only in genotypically diverse populations. □

Received 17 October 1989; accepted 19 February 1990.

- Levin, D. *Am. Nat.* **109**, 437–451 (1975).
- Jaenike, J. *Evol. Theory* **3**, 191–194 (1978).
- Glesener, R. R. & Tilman, D. *Am. Nat.* **112**, 659–673 (1978).
- Bremermann, H. J. *J. theor. Biol.* **87**, 641–702 (1980).
- Hamilton, W. D. *Oikos* **35**, 282–290 (1980).
- Lloyd, D. G. in *Evolutionary Biology* Vol. 13 (eds Hecht, M. K., Steere, W. C. & Wallace, B.) 69–111 (Plenum, New York, 1980).
- Bell, G. *The Masterpiece of Nature* (University of California Press, Berkeley, 1982).
- Antonovics, J. & Ellstrand, N. C. *Evolution* **38**, 103–115 (1984).
- Ellstrand, N. C. & Antonovics, J. *Evolution* **39**, 657–666 (1985).
- Schmitt, J. & Antonovics, J. *Evolution* **40**, 830–836 (1986).
- Parker, M. A. *Am. J. Bot.* **74**, 1758–1763 (1987).
- Burt, A. & Bell, G. *Nature* **326**, 803–805 (1987).
- Lively, C. M. *Nature* **328**, 519–521 (1987).
- Kelley, S. E., Antonovics, J. & Schmitt, J. *Nature* **331**, 714–716 (1988).
- Parker, M. A. *Evolution* **43**, 540–547 (1989).
- Lively, C. M. *Evolution* **43**, 1663–1671 (1989).
- Williams, G. C. *Sex and Evolution* (Princeton University Press, 1975).
- Maynard Smith, J. *The Evolution of Sex* (Cambridge University Press, 1978).
- Haldane, J. B. S. *La Ricerca Scientifica* (Suppl.) **19**, 68–76 (1949).
- Clarke, B. in *Genetic Aspects of Host-Parasite Relationships* Vol. 14 (eds Taylor, A. E. R. & Muller, R.) 87–103 (Blackwell Scientific, Oxford, 1976).
- Schultz, R. J. *Am. Nat.* **103**, 605–619 (1967).
- Vrijenhoek, R. C. *Science* **199**, 549–552 (1978).
- Moore, W. S. *Copeia* **1977**, 213–223 (1977).
- Cheng, T. C. *General Parasitology* 2nd edn (Academic, Orlando, Florida, 1986).
- Wakelin, D. in *Ecology and Genetics of Host-Parasite Interactions* (eds Rollinson, D. & Anderson, R. M.) 39–54 (Academic, London, 1985).
- Vrijenhoek, R. C. in *Speciation and its Consequences* (eds Otte, D. & Endler, J. A.) 386–400 (Sinauer, Sunderland, Massachusetts, 1989).
- Vrijenhoek, R. C. & Lerman, S. *Evolution* **36**, 768–776 (1982).
- Apple, J. L. in *Plant Disease* Vol. 1 (eds Harsfall, J. G. & Cowling, E. B.) 79–101 (Academic, New York, 1977).
- Freund, R. J. & Littell, R. C. *SAS for Linear Models* (SAS Institute, Cary, North Carolina, 1981).
- Sokal, R. R. & Rohlf, F. J. *Biometry* 2nd edn (Freeman, San Francisco, 1981).

ACKNOWLEDGEMENTS. We thank P. Smouse, L. Delph, O. Gaggiotti, S. Weeks, J. Hey, J. Quattro and M. McKone for helpful comments and the New Jersey Agricultural Experiment Station for support.

A cholecystokinin-like hormone activates a feeding-related neural circuit in lobster

G. G. Turrigiano & A. I. Selverston

Department of Biology, B-022, University of California at San Diego, La Jolla, California 92093, USA

THE peptide hormone cholecystokinin (CCK) contributes to the production of feeding-related behaviour in mammals, but the mechanism by which it exerts its effects remains unclear^{1–6}. The gastric mill neural circuit of lobster is an experimentally accessible model system for studying the hormonal control of feeding-related behaviour^{7,8}. Composed of 11 identified neurons, this circuit produces rhythmic movement of teeth within the stomach⁹. We have previously shown that the gastric mill motor pattern can be modulated by a cholecystokinin-like peptide *in vitro*¹⁰. We report here that (1) after feeding, levels of CCK-like peptide in haemolymph increase with the activation of the gastric mill, (2) injections of CCK activate the gastric mill, and (3) a specific CCK antagonist inhibits feeding-induced gastric mill activity. This neatly demonstrates a causal link between *in vivo* release of a peptide hormone and activation of a neural circuit.

In several vertebrates and invertebrates, blood levels of CCK-like peptide (CCKLP) rise after feeding^{11–14}. We compared changes in blood CCKLP levels in the lobster *Panulirus interruptus* with the activation of the gastric mill (GM) following feeding. GM activity was monitored in freely moving lobsters by recording the activity of the gm1 muscle, which protracts the medial tooth of the GM. In individual experiments, the GM

began cycling immediately after feeding, cycled continuously for 2–6 h, then the burst strength and frequency declined and activity became intermittent; this intermittent activity could continue for up to 48 h (Fig. 1a). On average, the GM reached its peak frequency during the first hour after feeding; the average frequency then declined over the next 3 h, but remained above base line for the next 4 h (Fig. 1b).

CCK-like peptide is found in lobster neurohaemal organs and haemolymph (G.G.T. and A.I.S., manuscript submitted). Here we use a radioimmunoassay to monitor the levels of CCKLP in haemolymph before and after feeding. There was a roughly fourfold increase in the levels of CCKLP in the first hour after feeding, peaking at an average of $(1.6 \pm 0.5) \times 10^{-9}$ M, and levels

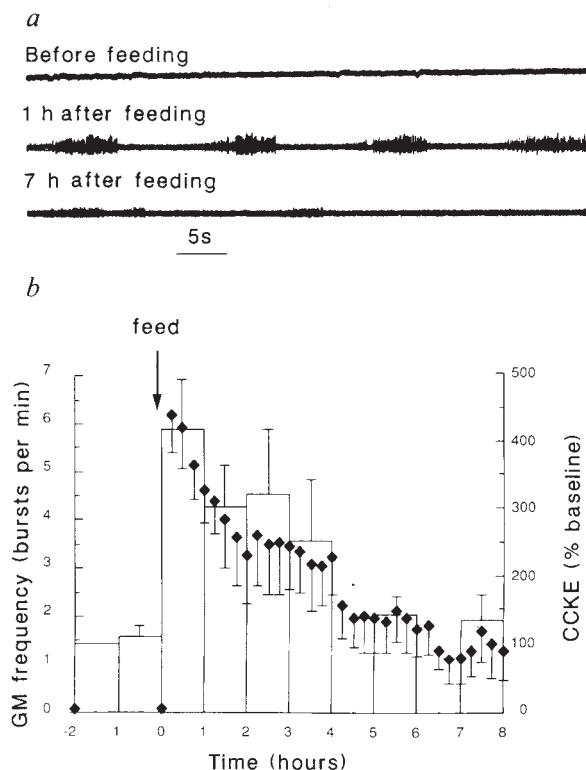


FIG. 1 Activity of the GM, and levels of CCKLP in haemolymph, before and after feeding. a, Gm1 muscle activity in one animal before feeding, at 1 h after feeding, and at 7 h after feeding. Before feeding there is no GM activity. At 1 h, the GM is cycling regularly, with a period of ~10 s. At 7 h the GM is cycling intermittently, and the individual muscle bursts are weaker than at 1 h. b, Average GM activity and average CCKLP levels in the haemolymph before and after feeding. Diamonds, average GM frequency, mean $\pm$ s.e.m. ($n=6$). Histograms, average CCKLP levels in haemolymph, mean $\pm$ s.e.m. ($n=7$); levels were significantly raised for the first 2 h after feeding ($P < 0.01$, Student's *t*-test).

METHODS. Myograms: Lobsters were starved for 1–3 weeks, kept on an artificial light-dark cycle, and fed ~10 g smelt or squid in the dark. Myograms were obtained from unrestrained animals by implanting teflon-coated silver wire electrodes into the gm1 muscle; electrode placement was verified at the end of the experiment³⁰. Animals recovered 2–7 days before initiation of experiments. Signals were differentially amplified, recorded on videocassette, and displayed with a Gould ES1000 electrostatic recorder. Generally 16 h of activity were recorded; activity was then monitored intermittently for up to 36 h. Muscle spikes were converted into constant-voltage pulses using a Schmidt trigger; pulses were digitized and analysed using an IBM PC. Bursts were defined as more than 20 spikes less than 1 s apart. Data were averaged over 15-min time blocks; values are mean $\pm$ s.e.m. CCKLP levels: Lobsters were maintained and fed as above. Aliquots (1 ml) of haemolymph were withdrawn from the ventral artery, extracted with ethanol, microfuged, the supernatants dried in a Speed-Vac concentrator, then reconstituted in buffer and assayed in duplicate by radioimmunoassay as described¹⁰; values are expressed as CCK molar equivalents (CCKE). Several haemolymph samples were obtained before feeding, and at ~1 h intervals thereafter. Data were normalized to first baseline value, and averaged over 1 h time blocks.

remained elevated for 4 h. This correlates well with the time food remains in the stomach¹⁵, and the time course of activation of the GM (Fig. 1b).

Injections of CCK in mammals produce changes in gut motility and satiety^{4,16,17}. Injections of 500 μ g CCK into lobster, which raised the haemolymph to between 10^{-7} and 10^{-6} M, initiated GM cycling in 9 out of 10 animals (Fig. 2); this concentration is higher than endogenous CCKLP after feeding. Preliminary data from reverse-phase analysis of CCKLPs in haemolymph and in the neurohaemal pericardial organs has revealed several species of CCKLP, none of which co-elutes with CCK, CCK8SO₄, the biologically active form of the peptide, or gastrin 17 (G.G.T and A.I.S., manuscript in preparation), suggesting that the endogenous peptide is not CCK. The endogenous peptide may be more potent than CCK at lower concentrations¹⁰. The two invertebrate CCKLPs sequenced to date are distinct from mammalian CCK^{18,19}. The GM activity stimulated by CCK (Fig. 2a) was not identical to the activity produced by feeding (Fig. 1a), but was generally slower. This was also true for the occasionally observed spontaneous GM activity in starved animals, and for GM activity produced by injection of the peptides proctolin or red pigment concentrating hormone (RPCH). These observations suggest that GM activity peaks only when there is food in the stomach. Stomach distention may supply several sources of excitation through activation of stretch or chemoreceptors.

Proglumide, a specific competitive antagonist of CCK²⁰⁻²², blocks the behavioural effects of CCK in mammals, including the induction of satiety^{5,23,24}. We tested whether proglumide could inhibit GM activity after feeding. Injections of proglumide (500 μ g) following feeding produced a long-lasting inhibition of the GM, whereas there were only short-lived interruptions after injections of vehicle (Fig. 3a). Proglumide inhibited GM activity for an average of 122 ± 22 min, which is significantly longer than the effect of vehicle alone (11 ± 5 min) (Fig. 3b);

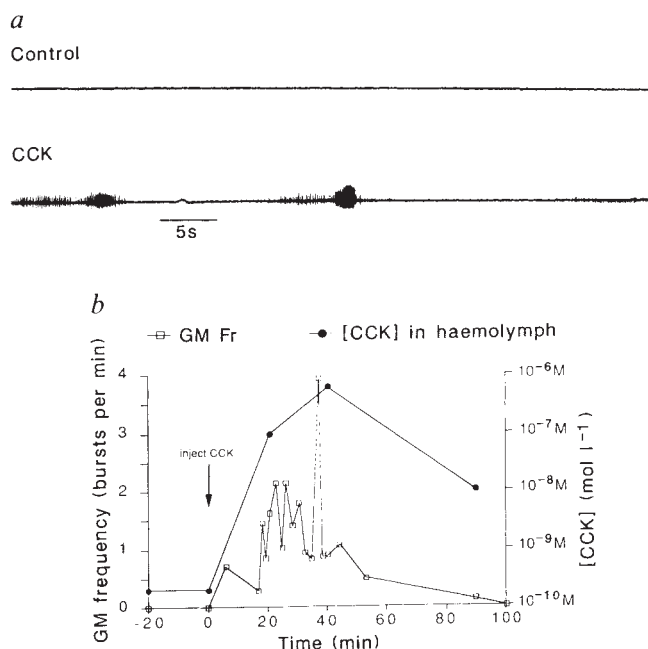


FIG. 2 GM activity before and after injections of CCK. *a*, Gm1 activity before injection of CCK (control), and following the injection of 500 μ g CCK8SO₄ into the ventral artery. *b*, Injection of CCK raised the haemolymph CCK levels to between 10^{-7} and 10^{-6} M, and activated the GM for ~60 min. Similar results were obtained in 9 out of 10 experiments; average latency to activation was 4.4 ± 1.1 minutes, average peak frequency (averaged over five minute periods) was 1.8 ± 0.4 bursts per min, and average duration was 40 ± 3.2 min. The GM usually fired bursts in bouts of 2 to 6 bursts, separated by periods of no activity. CCK concentration and GM activity were measured and analysed as described in the legend to Fig. 1.

results were independent of the order of injection. Minor disturbances of the animal, such as turning on a light transiently, caused brief interruptions in GM activity, comparable to those produced by vehicle injections (Fig. 3a). Injections of proglumide more than 2 h after the start of feeding had little effect on the GM (Fig. 3a, b), indicating that a CCKLP is probably necessary for early but not late GM activity. Two possible reasons for this are (1) that different mechanisms are responsible for GM activity after several hours, or (2) that longer exposures to CCKLP stimulate a second messenger system, the effects of which outlast the presence of the hormone.

To determine the specificity of proglumide in lobster, we investigated its ability to block the effects of CCK; we also tested two other peptides known to modulate the GM, proctolin²⁵ and RPCH²⁶. Proglumide prevented CCK from initiating GM activity when injected into the haemolymph (Fig. 4a) but it had no effect on the ability of RPCH to activate the GM (Fig. 4b). Proglumide also blocked the *in vitro* effects of CCK on the GM, but did not affect its modulation by proctolin or RPCH.

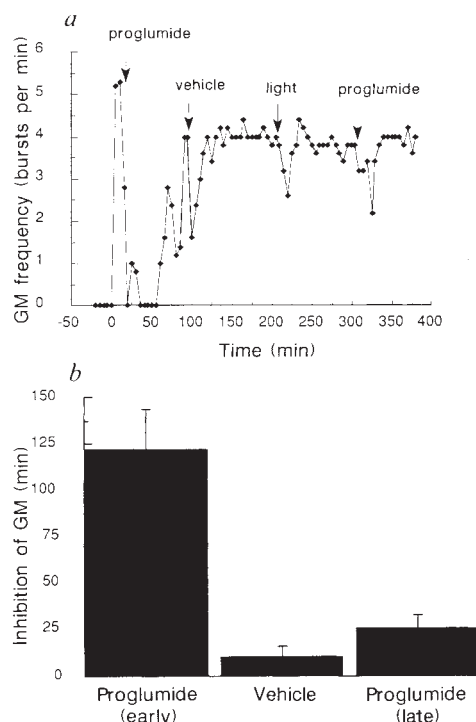


FIG. 3 The effect of the CCK antagonist proglumide on feeding-induced GM activity. *a*, Proglumide injected into the dorsal heart sinus soon after feeding produced a 70 min inhibition of GM activity; injection of vehicle, turning on the light, and injection of proglumide late in the experiment produced minor interruptions. *b*, Proglumide injections within the first 2 h after feeding produced a mean inhibition of 122 ± 22 min ($n=7$, mean $\pm$ s.e.m.). Injections of vehicle (1 ml saline + 5 μ l dimethyl sulphoxide) either before ($n=4$) or after ($n=3$) injections of proglumide, produced a mean inhibition of GM activity of 11 ± 5 min. Proglumide response was significantly different from that to the vehicle: $P < 0.001$, paired *t*-test. Injections of proglumide more than 2 h after the initiation of feeding produced a mean inhibition of 26 ± 7 min ($n=3$), which is not significantly different from vehicle, paired *t*-test.

METHODS. GM activity was monitored and analysed as described in the legend to Fig. 1. Drugs were injected by a catheter implanted in the dorsal heart sinus. Inhibition of GM activity was defined as the amount of time after injection in which the GM frequency was below 50% of its pre-injection value; once the frequency rose above this value for over 10 min, the effect was considered to be over. Proglumide was dissolved in dimethyl sulphoxide (5 mg per 50 μ l) immediately before use; 5 μ l was then diluted into 1 ml saline and injected. Normal GM cycling is virtually uninterrupted during the first 2 h after feeding; on average there are 10 ± 4.5 min during which the GM frequency falls below half of the average frequency of the first half hour ($n=6$).

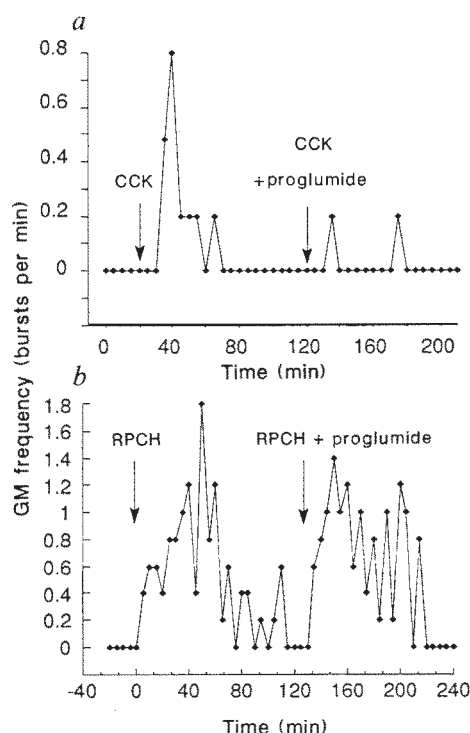


FIG. 4 The specificity of proglumide in lobster. *a*, Proglumide blocks the effects of CCK *in vivo*. Injection of CCK (500 μ g) activates the GM; injection of CCK (500 μ g) + proglumide (1 mg) has little or no effect. *b*, The ability of RPCH to activate the GM is unaffected by proglumide; injecting RPCH (200 μ g) or RPCH (200 μ g) + proglumide (1 mg) has the same effect. Controls were performed at least twice with similar results.

These data indicate that proglumide acts specifically to antagonize the effects of CCKLPs in lobster.

We infer that a CCKLP is released into the haemolymph after feeding, where it activates the GM. Because CCK antagonists inhibit the GM, a CCKLP must be necessary for feeding-induced GM activity, but as CCK injections cannot exactly reproduce the effect on GM following feeding, it is probably not sufficient. The GM receives a complex array of inputs, including several modulatory inputs and proprioceptive feedback^{10,25-29}. Besides its direct excitatory effects on the GM¹⁰, circulating CCKLP could act to prime the GM so that it can respond to these other inputs. Alternatively, circulating CCKLP may be responsible for directly activating other inputs which together induce vigorous GM activity.

We have previously identified fibres that can release CCKLP directly into the stomatogastric ganglion which contains the GM neural circuit¹⁰. It is unclear why the GM should receive both a hormonal and a modulatory CCK-like input. The two inputs may function together to raise CCKLP levels in the ganglion sufficiently to activate the GM. Alternatively, the two inputs may be independent, exerting different effects because of differences in concentration, in the number of targets, or as a result of co-release of several substances by the modulatory fibres.

The role of CCK in producing feeding-related behaviour in vertebrates is well documented. The similar role of CCKLP in lobster indicates a remarkable degree of conservation of function in this family of peptides. The identification of a specific antagonist for CCKLPs in lobster and the accessibility of the GM neural circuit, will allow us to resolve the cellular mechanisms by which this hormone exerts its behavioural effects to a level of detail difficult to attain in more complex systems. □

3. Della-Fera, M. A. & Baile, C. A. *Science* **206**, 471-473 (1979).
4. Sakatani, N. *et al. Peptides* **8**, 651-656 (1987).
5. Inui, A. *et al. Brain Res.* **417**, 355-359 (1987).
6. Dourish, C. T., Rycroft, W. & Iversen, S. D. *Science* **245**, 1509-1511 (1989).
7. Marder, E. *Nature* **335**, 296-297 (1988).
8. Bicker, G. & Menzel, R. *Nature* **337**, 33-39 (1989).
9. Selverston, A. I. & Moulins, M. (eds) *The Crustacean Stomatogastric System* (Springer, Berlin, 1987).
10. Turrigiano, G. G. & Selverston, A. I. *J. Neurosci.* **9**, 2486-2501 (1989).
11. Izzo, R. S., Brugge, W. R. & Praisman, N. *Reg. Pep.* **9**, 173-185 (1984).
12. Liddle, R. A., Goldfine, I. D. & Williams, J. A. *Gastroenterology* **87**, 542-549 (1984).
13. Linden, A., Uvnas-Moberg, K., Forsberg, G., Bednar, I. & Sodersten, P. *J. Endocrinol.* **121**, 59-65 (1989).
14. Van Wormhoudt, A., Favrel, P. & Guillaume, J. *J. comp. Physiol. B* **159**, 169-273 (1989).
15. Fleischer, A. G. *J. comp. Physiol.* **141**, 363-368 (1981).
16. Green, T., Dimaline, R., Peikin, S. & Dockray, G. J. *Am. J. Physiol.* **255**, G685-G689 (1988).
17. Collins, S. M. & Weingarten, H. P. *Can. J. Physiol. Pharmacol.* **64**, suppl., 54 (1986).
18. Nachman, R. J., Holman, G. M., Haddon, W. F. & Ling, N. *Science* **234**, 71-73 (1986).
19. Nichols, R. S., Schneuwly, A. & Dixon, J. E. *J. Biol. Chem.* **263**, 12167-12170 (1988).
20. Hahne, W. F., Jensen, R. T., Lemp, G. F. & Gardener, J. D. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6304-6308 (1981).
21. Gardner, J. D. & Jensen, R. T. *Am. J. Physiol.* **246**, G471-G476 (1984).
22. Chiodo, L. A. & Bunney, B. S. *Science* **219**, 1449-1451 (1983).
23. Katsura, G., Hsiao, S. & Itoh, S. *Peptides* **5**, 529-534 (1984).
24. Crawley, J. N., Stivers, J. A., Hommer, D. W., Skirboll, L. R. & Paul, S. M. *J. Pharmac. exp. Ther.* **236**, 320-330 (1986).
25. Heinzel, H.-G. & Selverston, A. I. *J. Neurophysiol.* **59**, 566-585 (1988).
26. Dickenson, P. S. & Marder, E. *J. Neurophysiol.* **61**, 833-844 (1989).
27. Dickenson, P. S., Nagy, F. & Moulins, M. *J. exp. Biol.* **136**, 53-87 (1988).
28. Katz, P. S. & Harris-Warick, R. M. *J. Neurophysiol.* **62**, 571-581 (1989).
29. Dickenson, P. S. & Marder, E. *J. Neurophysiol.* **61**, 833-844 (1989).
30. Rezer, E. & Moulins, M. *J. comp. Physiol.* **153**, 17-28 (1983).

ACKNOWLEDGEMENTS. We thank L. Ogden for technical assistance, and R. M. Harris-Warick, N. Herterich, B. Kristan, S. R. Lockery, E. Marder, S. Nelson and G. Wittenberg for their comments. This work was supported by a grant from the NIH.

Activin is a nerve cell survival molecule

D. Schubert, H. Kimura, M. LaCorbiere, J. Vaughan, D. Karr & W. H. Fischer

The Salk Institute, 10010 N. Torrey Pines Road, La Jolla, California 92037, USA

THE structures of five neurotrophic molecules have so far been published. Nerve growth factor¹, fibroblast growth factor^{2,3} and purpurin⁴, have been identified as nerve-cell survival molecules. More recently, brain-derived neurotrophic factor (BDNF) and ciliary neurotrophic factor have been cloned and sequenced^{5,6}. As all these proteins stimulate the survival of ciliary or sensory neurons, a new cell survival assay is required if novel neurotrophic molecules are to be discovered. P19 teratoma cells differentiate to nerve-like cells in the presence of 5×10^{-7} M retinoic acid (RA)^{7,8}. But when P19 cells are plated in N_2 synthetic medium⁹ without being exposed to RA, they die within 48 h. In an attempt to identify a molecule(s) that can substitute for RA in promoting P19 survival, we assayed serum-free growth-conditioned media for their ability to promote P19 survival. One cell line from the rat eye secreted a molecule that promoted the survival of P19 cells and some types of nerve cell. We identified this molecule as activin, better known for its role in hormone secretion.

When the serum-free growth-conditioned media of many clonal cell lines from the eye and central nervous system¹⁰ were assayed for their ability to promote the survival of P19 cells, one cell line could do so in a dose-dependent manner. The clonal cell line, designated R33, was obtained from an eye tumour induced by nickel subsulphide¹¹. The exact phenotype of R33 is unknown. Figure 1a shows that at concentrations $>10\%$, medium conditioned by R33 completely maintained P19 viability for 3 days in N_2 medium. The resulting population of cells is 83% A_2B_5 positive and morphologically homogeneous. Time-lapse photography shows that they do not divide. The phenotype of the surviving cells is not known, nor is the response of the P19 clone used in these experiments likely to be typical of all embryonal carcinoma cell lines. As a negative control, medium conditioned by PC12 had no effect. The R33 survival

Received 29 November 1989; accepted 21 February 1990.

1. Smith, G. P. & Gibbs, J. *Ann. N.Y. Acad. Sci.* **448**, 417-423 (1985).

2. Gibbs, J., Young, R. C. & Smith, G. P. *J. comp. Physiol. Psychol.* **84**, 488-495 (1973).

activity bound to heparin Sepharose and eluted at 0.6 M sodium chloride.

When the affinity column eluant was applied to a reverse phase column, the survival activity eluted at about 40% acetonitrile (Fig. 2). On SDS acrylamide gels the active HPLC fractions contained a major band of relative molecular mass 30,000 (M_r 30K) which, when eluted from the acrylamide gels, contained the cell-survival activity (Fig. 3a, lane 7, b). On

TABLE 1 Amino-acid sequences of R33 survival protein

A	Gln-Phe-Phe-Val-Ser-Phe-Lys
B	Gly-X-Ser-Pro-Phe-Ala-Asn-Leu-Lys
C	Ser-Met-Leu-Tyr-Tyr-Asp-Asp-Gly-Gln-Asn-Ile-Ile-Lys

Amino-acid sequences obtained by Edman degradation of proteolytic fragments of R33 survival protein. Protein was electrophoretically transferred to nitrocellulose, stained with amido black 10B, and protein bands excised. Nitrocellulose pieces were treated with polyvinyl pyrrolidone and washed with water. Trypsin (1 μ g, Sequencing Grade, BMB) was added in 0.1 M TES, pH 8, containing 5% acetonitrile and 2 mM calcium acetate. Proteolysis was allowed to proceed for 16 h at 37 °C. Peptide fragments released into the supernatant were separated on a Vydac C18 HPLC column (2 $\times$ 150 mm; pore size, 300 Å; particle size, 5 μ m). Sequence analysis was performed on an ABI 470A Protein Sequencer equipped with an on-line ABI 120A PTH Analyzer. Yields for PTH amino-acid derivatives were between 40 and 2 pmol. Sequences are identical to activin β_A (résidues 16–22, 71–79, and 91–103, respectively¹⁴). X, unidentified residue. Activin β_B is considerably different in these regions.

reduction with 5% mercaptoethanol, this protein generated a band at M_r 15K which was biologically inactive (Fig. 3a, lane 5).

To determine a partial amino-acid sequence of the suspected survival factor, the most active HPLC fraction was run on an SDS gel, electrophoretically transferred to nitrocellulose, and the active band excised from the nitrocellulose after staining with amido black. The protein was digested *in situ* with trypsin^{12,13} and the eluted peptides separated on a C-18 reverse phase column. Three peptide sequences were obtained which were identical to the β_A chain of rat activin A (ref. 14) (Table 1). To confirm this result, the native and reduced forms of the R33 protein and authentic activin A were run together on an SDS acrylamide gel and immunoblotted with an antiserum against a synthetic peptide fragment of activin β_A specific for the β_A chain^{15,16}. Figure 3 (lanes 1–4) shows that the R33 protein and activin A co-migrated in both reduced and unreduced forms. The R33 protein did not react with antiserum specific for the β_B -chain^{15,16}.

To verify the functional identity between the protein derived from the eye cell line and activin A, highly purified activin A prepared from porcine follicular fluid¹⁷ was tested in the P19 survival assay. Figure 1b shows that follicular activin is maximally active at a concentration of about 20 picomolar. By contrast, inhibin A (M_r 32K), purified from porcine follicular fluid using immunoaffinity chromatography¹⁵ was inactive. Competition assays in which increasing amounts of inhibin A were used to block the half-maximal response of activin A on P19 cells, showed that there was no inhibition at concentrations up to 100-fold molar excess (data not shown).

Because the serum-free medium conditioned by R33 and the activin-like molecule derived from it promote the survival of the P19 cell line, which can give rise to nerve among other cell types, we investigated whether activin A could stimulate nerve cell survival in other assay systems—the chick neural retina and ciliary ganglion, and a central nervous system (CNS) nerve cell line. Activin A promoted the survival of the B50 nerve cell line derived from the rat CNS (ref. 10) (Fig. 1c). Figure 1d shows the survival of 10-day embryonic neural retina cells in the presence of activin. Inhibin A was inactive in promoting the survival of the neural retina cells, but promoted the survival of the CNS nerve cell line at relatively high concentrations (Fig. 1c). The half-maximal concentration of activin A for B50 survival is about 20 picomolar, and for inhibin A, 60 picomolar. Neither activin A nor inhibin A increased the survival of ciliary ganglion cells in culture using the standard assay¹⁸, nor did they promote the survival of several glial, myoblast or fibroblast cell lines (data not shown).

Activin and inhibin are disulphide-linked dimeric proteins isolated on the basis of their ability to modulate the release of

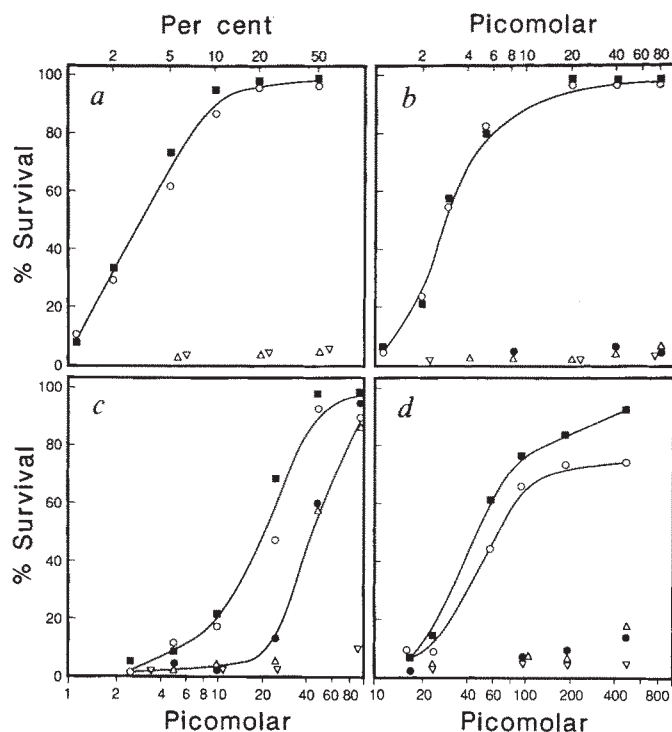


FIG. 1 Cell survival assays. a, P19 survival. Exponentially dividing P19S1801A1 cells⁷ were obtained from Dr J. Levine. The P19 clone P19S1801A1 (ref. 7) used in these experiments was not subcloned. The cells were, however, passaged several times in Dr Levine's laboratory and our own, and may not be representative of all embryonal carcinoma cell lines or of other P19 clones. They were dissociated with 10% pancreatin (Gibco), washed once with 10% FCS in DMEM medium, twice with N_2 medium, and plated at 5×10^4 cells per 35-mm uncoated tissue culture dish. The dissociation of cells with pancreatin is critical to the success of these experiments. Increasing amounts of serum-free growth-conditioned medium from R33 or PC12 cells were added between 2 and 30% of the volume. The serum-free growth-conditioned medium was prepared by washing semi-confluent cultures twice with serum-free medium and then placing the cells in 3 ml per 100-mm plate of serum-free medium for 2 days. This conditioned medium was then centrifuged at 2,000 r.p.m. to remove cells, and at 15,000 r.p.m. to remove debris, and used for assays and protein purification. ■ and ○, R33-conditioned medium, duplicates; Δ and ∇, PC12-conditioned medium, duplicates. No addition, <5% survival. Survival was determined visually by trypan blue exclusion or with the metabolic dye, MTT (ref. 29), and plotted as the percentage of total input cell number that is alive at 3 days. All viable cells were scored irrespective of apparent phenotype. b, Effect of activin on P19 survival. Increasing amounts of activin A (○ and ■) or porcine inhibin A (Δ and ∇) or buffer (●) were added to P19 cells, and viability determined as described in A. c, B50 nerve cell line. Exponentially dividing B50 cells were dissociated and plated in 35-mm tissue culture dishes as described for P19, except that cells were plated into serum-free DMEM. Increasing amounts of activin A (○ and ■) or inhibin A (● and Δ) or buffer (∇) were added and percentage of input cells alive determined after 3 days. d, Neural retina survival. Ten-day embryonic chick neural retina were dissociated with trypsin and plated at 2×10^4 cells per 35-mm tissue culture plate pre-coated with 1 μ g mouse laminin (Sigma) in serum-free DMEM (ref. 4). The number of viable cells was determined after 5 days, as above. There was no cell division under these conditions as determined by time-lapse photography. ■ and ○, activin A, separate experiments; ● and Δ, inhibin A; ∇, buffer.

FIG. 2 Reverse phase purification of survival activity. One litre of serum-free growth-conditioned medium of R33 was initially adsorbed to a heparin-Sepharose (Pharmacia) affinity column and eluted with 0.6 M NaCl in 0.01 M Tris, pH 7.4. The fraction biologically active in the P19 survival assay (Fig. 1) was then acidified and pumped onto a 250 × 4.6 mm SynChropak RP-4 column, and the proteins eluted with a linear gradient of 20–70% acetonitrile over 3 h (solid straight line). Each fraction was assayed for cell survival activity (hatched area) and the optical density was continuously monitored at $\lambda = 212$ nm (wavy line).

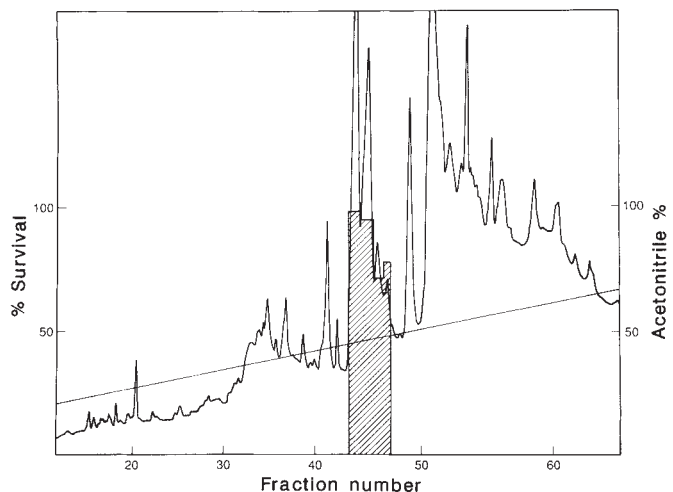
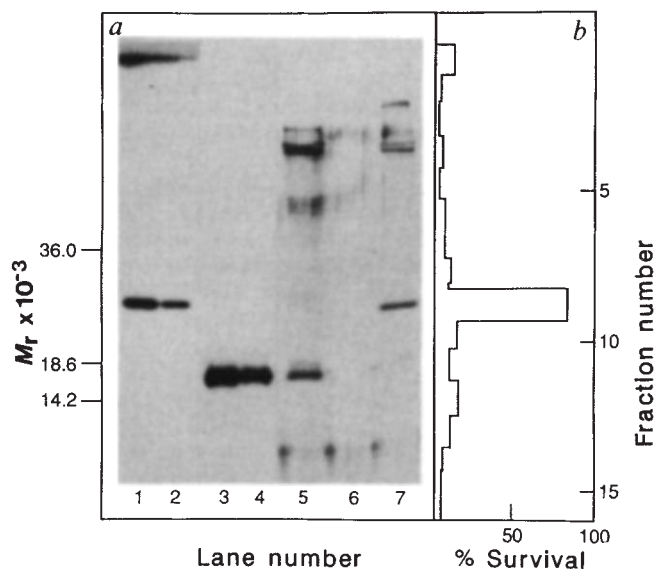


FIG. 3 Acrylamide gel electrophoresis of R33 protein and activin. All samples were run on 15% acrylamide gels containing 0.1% SDS. In *b* an identical sample in a contiguous lane was cut into 15 equal fractions, the proteins eluted by shaking at 4 °C overnight in 0.02% SDS, and the eluant assayed for biological activity on P19 cells as described in Fig. 1. In other cases the proteins were immunoblotted with affinity-purified rabbit anti-activin β_A (residues 81–113) which is specific for the β_A chain^{15,16}. *a*, Lane 1, immunoblot of R33 HPLC fraction 43 (used for sequencing), unreduced; lane 2, immunoblot of activin A, unreduced; lane 3, immunoblot of R33 HPLC fraction 43 reduced; lane 4, immunoblot of activin A reduced; lane 5, silver stain of HPLC fraction 43 reduced; lane 6, blank; lane 7, silver stain of HPLC fraction 43 unreduced. *b*, P19 cell survival of HPLC fraction 43 run in lane next to lane 7, eluted, and assayed for biological activity.

follicle-stimulating hormone (FSH) from pituitary cells^{17,19}. Inhibin is composed of two distantly related peptide chains, α and β , whereas activin is a β -chain homodimer. In addition, there are two closely related types of β -chain, called β_A and β_B . There are therefore two possible forms of inhibin, and three of activin. The activin secreted by R33 cells is a β_A chain homodimer (activin A), as the only sequences detected among the tryptic peptides of the biologically active molecule were β_A peptides, and the protein reacted strongly with an antiserum to β_A . Although activin was initially isolated by its ability to stimulate FSH secretion, it was recently shown that activin A is identical to another protein, erythroid differentiation factor, which causes the differentiation of Friend erythroleukaemia cells^{20,21}. Also, activin is weakly mitogenic for BALB/c 3T3 cells²², growth inhibitory for rat thymocytes²³, and is found in

a limited area of the adult brain²⁴. The P19 cells used to isolate activin have a high density of activin receptors²⁵.

The above results show that a cell line derived from the rat eye can synthesize and secrete activin A and that this protein promotes nerve cell survival. Activin could, therefore, have additional functions in the developing animal apart from its role in hormone secretion. Judging from nucleotide sequences, the activin β -chain has some structural homology to the product of a *Drosophila* decapentaplegic gene complex which has a role in early development²⁶ and the VG1 gene which may mediate embryonic mesoderm formation²⁷. As the biology of teratoma cells may reflect early events in mammalian development²⁸, activin could function in a manner similar to VG1 or decapentaplegic in the mammalian embryo as well as having a neurotrophic role in the more mature nervous system. □

Received 29 January; accepted 2 March 1990.

- Levi-Montalcini, R. *Prog. Brain Res.* **45**, 235–258 (1976).
- Baird, A. & Böhlen, P. in *Peptide Growth Factors and their Receptors* (eds Sporn, M. B. & Roberts, A.) 197–225 (Springer-Verlag, Heidelberg) (1990).
- Schubert, D., Ling, N. & Baird, A. *J. Cell Biol.* **104**, 635–643 (1987).
- Berman, P. *et al. Cell* **51**, 135–142 (1987).
- Lin, L. F. *et al. Science* **246**, 1023–1025 (1989).
- Leibrock, J. *et al. Nature* **341**, 149–151 (1989).
- Edwards, M. K. & McBurney, M. W. *Dev. Biol.* **98**, 189–191 (1983).
- Levine, J. M. & Flynn, P. *J. Neurosci.* **6**, 3374–3381 (1986).
- Bottenstein, J. & Sato, G. *Proc. natn. Acad. Sci. U.S.A.* **76**, 514–518 (1978).
- Schubert, D. *et al. Nature* **249**, 224–227 (1974).
- Albert, D. M. *et al. Invest. Ophthalmol. Vis. Sci.* **22**, 768–782 (1982).
- Aebbersold, R. H. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 6970–6974 (1987).
- Gonzalez, G. A. *et al. Nature* **337**, 749–752 (1989).
- Woodruff, T., Meunier, H., Jones, P. B., Hsueh, A. J. & Mayo, K. E. *Molec. Endocrinol.* **1**, 561–570 (1987).
- Vaughan, J. M. *et al. Meth. Enzym.* **168**, 588–617 (1989).
- Meunier, H. *et al. Molec. Endocrinol.* **2**, 1352–1363 (1988).

- Vale, W. *et al. Nature* **321**, 776–779 (1986).
- Collins, F. *Dev. Biol.* **65**, 50–57 (1978).
- Mason, A. J. *et al. Nature* **318**, 659–663 (1985).
- Eto, Y. *et al. Biochem. biophys. Res. Commun.* **142**, 1095–1103 (1987).
- Murata, M., Eto, Y., Shibai, H., Sakai, M. & Muramatsu, M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2434–2438 (1988).
- Kojima, I. & Ogata, E. *Biochem. biophys. Res. Commun.* **159**, 1107–1113 (1989).
- Hedger, M. P., Drummond, A. E., Robertson, D. M., Risbridger, G. P. & de Kretser, D. M. *Molec. cell. Endocrinol.* **61**, 133–138 (1989).
- Sawchenko, P. E. *et al. Nature* **334**, 615–617 (1988).
- Kondo, S. *et al. Biochem. biophys. Res. Commun.* **161**, 1267–1271 (1989).
- Padgett, R. W., Johnston, D. S. & Gelbart, W. M. *Nature* **325**, 81–83 (1987).
- Weeks, D. L. & Melton, D. A. *Cell* **51**, 861–867 (1987).
- Martin, G. R. *Science* **209**, 768–776 (1980).
- Pettman, B., Manthorpe, M., Powell, J. A. & Varon, S. *J. Neurosci.* **8**, 3524–3532 (1988).

ACKNOWLEDGEMENTS. We thank Drs Pamela Maher and Wylie Vale for discussions, Dr Joel Levine for the cells, and Belle Wamsley for preparation of the manuscript. We will provide the P19 clone used in these experiments on request. This work was supported by the National Eye Institute and National Institutes of Health.

Role of myelin Po protein as a homophilic adhesion molecule

Marie T. Filbin*, Frank S. Walsh†, Bruce D. Trapp*,
John A. Pizzey† & Gihan I. Tennekoon*

* Department of Neurology, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

† Department of Experimental Pathology, UMDS, Guy's Hospital, London SE1 9RT, UK

PERIPHERAL nervous system myelin is an extension of the Schwann cell's plasma membrane that tightly enwraps axons in many layers and permits nerve impulses to be rapidly conducted. It is not known how these multiple membrane layers are held together in this compact form. Here we present evidence supporting the hypothesis that the extracellular leaflets of myelin¹⁻⁴ are held together by the most abundant protein of myelin of the peripheral nervous system, Po, by homophilic interaction of its extracellular domains. Transfected Chinese hamster ovary cells expressing Po protein adhere to each other in suspension, to form large aggregates, whereas cells that are identical but which do not express Po do not. We also show that this aggregation is mediated by homophilic binding between Po-expressing cells and that the apposing plasma membranes of these cells specifically form desmosomes, whereas control transfected cells do not. As the only difference between the two cell populations is the expression of Po, this protein is apparently responsible for the changes in morphology and adhesion in the cells that express it. The idea that Po is a homophilic adhesion molecule is supported by its inclusion in the immunoglobulin supergene family⁵, all members of which are involved in recognition and/or adhesion⁶.

We have shown previously that CHO cells transfected with a plasmid containing Po complementary DNA and selected for gene amplification by the dihydrofolate reductase (DHFR)/methotrexate strategy express high levels of Po glycoprotein, a portion of which reaches the cell's surface⁷. Hence, these cells are ideal for studying the adhesive properties of Po, especially as Schwann cells in culture express very little Po.

To assess the adhesive properties of these Po-expressing cells, a suspension of single cells was allowed to aggregate. Assay

TABLE 1 Intercellular collision efficiencies (α) of transfected cells

Cell	Conditions	Adhesiveness (α)
No Po (control)	+serum; +Ca ²⁺	3.6 ± 1.79
Po-expressing	+serum; +Ca ²⁺	7.75 ± 2.47
Po-expressing	-serum; -Ca ²⁺	8.09 ± 2.04

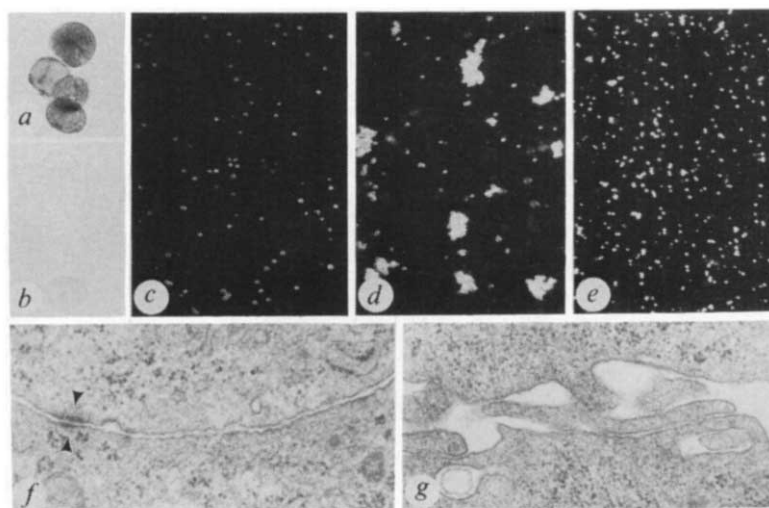
Aggregation was assayed as described for Fig. 1, except that α was determined in the absence of Ca²⁺ as follows: cells were washed in Ca²⁺-free MEM and finally resuspended in Ca²⁺-free MEM containing 10 mM HEPES and 5% Ficoll. Aggregation was not continued beyond 90 min as the percentage of cells viable after this time dropped dramatically in serum-free medium. The α value, the collision efficiency of adhesiveness of the cells, was calculated as follows. The probability that a collision between two particles results in adhesion is given by $\ln N_t/N_0 = -4G\phi\alpha t/\pi$, where N_t and N_0 are the total number of particles at time t and time zero respectively; G is the shear rate of the suspending medium and ϕ is the volume fraction of the particles in suspension. If the cells are allowed to reaggregate in a laminar flow of a constant shear rate (a couette viscometer), α can be calculated directly by measuring the change in particle number with time⁸.

conditions, particularly the shear forces, are carefully controlled, by allowing aggregation to take place in a couette viscometer⁸⁻¹¹; adhesiveness can then be quantitated and a relative value assigned to the interaction (see the legend to Table 1). Aliquots taken at intervals were examined under the microscope and counted for total particle number in a Coulter counter. Results were compared with those from control cells that had been transfected with the same plasmid without the Po cDNA insert. Using immunostaining, we see that Po is still present at the surface of these cells in suspension (Fig. 1a and b). After incubation at 37 °C in a couette viscometer for 70 min, large aggregates are apparent in the Po-expressing cell suspension (Fig. 1c and d), whereas cells transfected with control plasmid form only doublets and triplets in the same time (Fig. 1e). At higher magnification, cells in both populations are phase-bright and rounded after 70 min, which is a good indication of their viability even when they are in large clumps. It is unlikely therefore that the clumps result from nonspecific aggregation of dead cells. We conclude that Po expression dramatically increases the aggregation and adhesion of CHO cells.

Electron microscopy of cells incubated for 60 min confirmed the results from our light micrographs. After 60 min incubation,

FIG. 1 a and b, Single cell suspensions of Po-expressing (a) and control cells (b) immunostained for Po. Cells were fixed in 4% paraformaldehyde for 30 min at room temperature, treated with 0.25% Triton X-100 for 30 min at 4 °C, and immunostained with Po antiserum¹⁷ (diluted 1 in 250) by the avidin-biotin procedure, using diaminobenzidine as chromogen¹⁸. c, d and e, Representative micrographs of dissociated transfected CHO cells after different periods of aggregation.

METHODS. Po-expressing cells at time zero (c) and after 70 min aggregation (d); e, control cells after 70 min aggregation. Dihydrofolate reductase (DHFR) mutant CHO cells were transfected with plasmid containing Po cDNA and the DHFR gene. Characterization of Po expression by these cells has been described⁷. Controls were CHO cells transfected with a similar plasmid carrying no Po cDNA. Cells at 80–90% confluence were washed with PBS and incubated with 0.025% trypsin for 2–3 min at room temperature, washed twice in minimal essential medium (MEM) containing 10 mM HEPES, and finally resuspended in MEM containing 10% fetal calf serum, 5% Ficoll and 10 mM HEPES by passage through a 19-gauge syringe. Suspensions containing a minimum of 95% single cells were diluted to a final concentration of 0.5–1 × 10⁶ cells per ml and allowed to aggregate in a couette viscometer at 37 °C under controlled microenvironment conditions⁸. Aliquots were removed at intervals, examined under the microscope and their total particle number counted with a Coulter counter, details as described⁸⁻¹¹. f and g, Electron micrographs of Po-



expressing (f) and control (g) cells after 60 min incubation. Cells were fixed with 4% paraformaldehyde and 2% glutaraldehyde, processed to Epon by routine procedures and examined in a Hitachi H-600 electron microscope. Magnifications in a and b, ×500; in c–e, ×40; in f, ×51,000, and in g, ×43,400.

the plasma membranes of Po-expressing cells were flat and straight where they came into contact with each other. The space between them was uniformly narrow and these apposing membranes often formed desmosomes (Fig. 1*f*, arrowed). The plasma membrane of the control cells contained many small finger-like projections. In rare instances when two control cells were apposed, the space between them varied over a short distance and contact was usually through finger-like projections (Fig. 1*g*); no desmosomes were seen. These differences indicate that expression of Po in CHO cells was responsible for such specific membrane-membrane interactions.

We plotted this difference in aggregation as the total number of particles against time; a decrease in the number of particles is indicative of aggregate formation. Figure 2 shows that the total number of particles decreases to 20% in 70 min for the Po-expressing cells, but to only 60% for the control cells over the same time. These values agree with our microscopic observations (Fig. 1*d* and *e*); most of the Po-expressing cells are in large aggregates, some containing up to 40 cells, whereas the controls are mainly single cells, doublets or triplets. When a value is assigned to the adhesiveness, or α (as described in the legend to Table 1), for Po-expressing cells α is calculated as twice the log value for the controls (Table 1). The same value

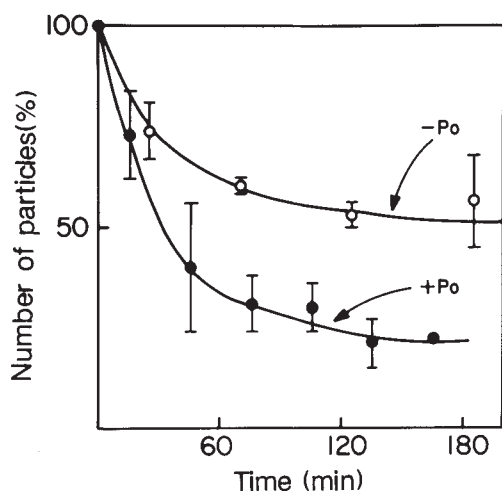


FIG. 2 Aggregation kinetics of dissociated CHO cells transfected with a plasmid containing Po cDNA (+Po) or with the same plasmid without Po cDNA (-Po). Points are the mean of at least three experiments, each in triplicate. Bars represent standard errors. Aggregation was assayed as described in the legend to Fig. 1. The total particle number was counted with a Coulter counter with the aperture open wide to allow passage of large cell aggregates.

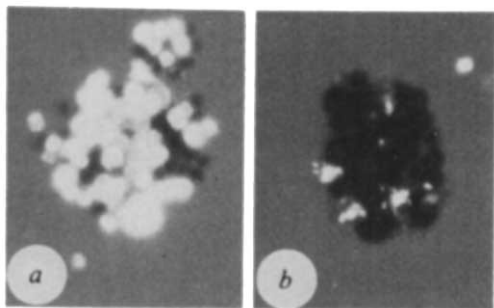


FIG. 3 Fluorescein-labelled Po-expressing or control cells adhering to unlabelled preformed aggregates. Cells expressing Po were allowed to aggregate for 45 min as described in the legend to Fig. 1. A single cell suspension of either Po-expressing or control cells labelled with CFDA⁸, was then added to the preformed aggregates and incubated for a further 45 min, after which aliquots were examined by both phase and fluorescence microscopy¹⁹. *a*, Po-expressing cells labelled with CFDA added to preformed aggregates; *b*, control cells labelled with CFDA added to preformed aggregates. Magnification, $\times 140$.

for α was obtained for two different clonal cell lines expressing equivalent amounts of Po, indicating that the adhesiveness of the CHO cells expressing Po is not an artefact of one particular cell line (results not shown). Furthermore, the increased adhesiveness of CHO cells induced by Po expression is independent of calcium (Table 1); this is not surprising as the adhesive properties of other members of the immunoglobulin supergene family are also unaffected by calcium¹².

We next determined whether the adhesive properties of Po-expressing cells resulted from a homophilic interaction with one another, or from a heterophilic interaction between Po and some other component of the CHO cell's plasma membrane. We mixed equal numbers of Po-expressing and control cells in suspension and allowed them to aggregate in a mixed cell-adhesion assay^{8,13}. One population was prelabelled with the dye 6-carboxyfluorescein diacetate (CFDA). After 20 min, aliquots were removed and for each experiment 50 four-cell aggregates were scored for Po-expressing cells or for control cells using phase and fluorescence microscopy and statistically evaluated as described¹³. If the cell types have no adhesive preference, aggregation will result from random heterophilic interaction, but Po-expressing cells adhering preferentially to each other will show non-random aggregation. In three experiments we found that a mixture of Po-expressing and control cells aggregated non-randomly ($P < 0.0001$), indicating that the interaction was homophilic for Po. We then allowed Po-expressing cells to form aggregates for 45 min, after which we added a single cell suspension of either CFDA-labelled Po-expressing cells or CFDA-labelled control cells and incubated the mixture for a further 45 min. As can be seen in Fig. 3, only the Po-expressing cells adhered to the preformed aggregates; control cells remained as single cells or doublets.

After statistical evaluation of our results, we conclude that Po behaves as a homophilic adhesion molecule, responsible for dramatically increasing the calcium-independent aggregation of CHO cells with respect to control cells transfected with a similar plasmid without a cDNA insert coding for Po. Adhesion seems to be homophilic and results in the formation of specific cell-cell junctions through a single immunoglobulin. Another protein expressed by myelin-forming V-like domain. Schwann cells which is also a member of the immunoglobulin superfamily, the myelin-associated glycoprotein MAG, has recently been described as an adhesion molecule, participating in both homophilic and heterophilic interactions^{14,15}. MAG is present in membranes of Schmidt-Lanterman incisures, paranodal loops, and outer and inner mesaxon, in which it possibly functions to maintain the periodicity of these membranes¹⁶. But as MAG is absent in compact myelin it is presumably not helping to keep the myelin membrane in its compact state. On the other hand, Po protein is clearly involved in the compaction of myelin. □

Received 4 January; accepted 15 February 1990.

- Kirschner, D. A. & Ganster, A. L. *Nature* **283**, 207-210 (1980).
- Braun, P. E. in *Myelin* (ed. Morell, P.) 97-116 (Plenum, New York, 1984).
- Lemke, G. & Axel, R. *Cell* **40**, 501-508 (1985).
- Trapp, B. J. *J. Cell Biol.* **107**, 675-685 (1988).
- Lai, C. et al. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4337-4341 (1987).
- Barclay, A. N., Johnson, P., McCaughan, G. W. & Williams, A. F. in *The T Cell Receptors* (ed. Mak, T. W.) 53-87 (Plenum, New York, 1988).
- Filbin, M. T. & Tennekoon, G. I. *J. Neurochem.* (in the press).
- Pizzey, J. A., Jones, G. E. & Walsh, F. S. *J. Cell Biol.* **107**, 2307-2317 (1988).
- Jones, G. E. & Witkowski, J. A. *J. neuro. Sci.* **43**, 465-470 (1979).
- Jones, G. E. & Witkowski, J. A. *J. Cell Sci.* **48**, 291-300 (1981).
- Pizzey, J. A. & Jones, G. E. *J. neuro. Sci.* **69**, 207-221 (1985).
- Rutishauser, U., Hoffman, S. & Edelman, G. M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 685-689 (1982).
- Siebert, F. & Roseman, S. *J. Cell Biol.* **90**, 55-62 (1981).
- Poltorak, M. et al. *J. Cell Biol.* **105**, 1893-1899 (1987).
- Johnson, P. W. et al. *Neuron* **33**, 377-385 (1989).
- Trapp, B. D. & Quarles, R. H. *J. Cell Biol.* **92**, 877-882 (1982).
- Trapp, B. C., McIntyre, L. J., Quarles, R. H., Sternberger, N. H. & Webster de F. H. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3552-3556 (1979).
- Stoll, G., Trapp, B. D. & Griffin, J. W. *J. Neurosci.* **9**, 2327-2335 (1989).
- Snow, P. M., Bieber, A. J. & Goodman, C. S. *Cell* **59**, 313-323 (1989).

ACKNOWLEDGEMENTS. We thank Dr Pamela Talalay for discussions, Peter Hauer for help with the photographs and Diane Emerson for typing the manuscript. This work was supported by grants from the NIH, NSF, Wellcome Trust and NATO.

Induction of CD8⁺ cytotoxic T cells by immunization with purified HIV-1 envelope protein in ISCOMs

Hidemi Takahashi^{*†}, Toshiyuki Takeshita^{*},
Bror Morein[‡], Scott Putney[§], Ronald N. Germain^{||}
& Jay A. Berzofsky^{*¶}

^{*} Molecular Immunogenetics and Vaccine Research Section,
Metabolism Branch, National Cancer Institute,
National Institutes of Health, Bethesda, Maryland 20892, USA

[‡] Department of Virology, National Veterinary Institute, Biomedical Center,
S-751 23 Uppsala, Sweden

[§] Repligen Corporation, Cambridge, Massachusetts 02139, USA

^{||} Lymphocyte Biology Section, Laboratory of Immunology,
National Institute of Allergy and Infectious Disease,
National Institutes of Health, Bethesda, Maryland 20892, USA

To reduce the risks of immunization with killed or live attenuated virus vaccines, it may be advantageous to use a pure, defined antigen that contains determinants for both humoral and cellular immunity. However, although most non-living intact protein preparations induce antibodies and CD4⁺ major histocompatibility complex (MHC) class II-restricted helper and/or cytotoxic T lymphocytes (CTL), they do not elicit CD8⁺ MHC class I restricted CTL^{1,2}. Indeed, with a few exceptions^{3,4}, it has not so far been possible to induce CD8⁺ CTL by immunizing with intact soluble proteins. We show here that a single subcutaneous immunization in mice with immunostimulating complexes containing either purified intact gp160 envelope glycoprotein of the human immunodeficiency virus (HIV)-1 or influenza haemagglutinin results in reproducible and long-lasting priming of HIV specific or influenza-specific CD8⁺, MHC class I restricted CTL.

Immunostimulating complexes (ISCOMs) are stable molecular structures, with a mean diameter of 35 nm, in which protein or peptide antigens are incorporated into a matrix of the adjuvant glycoside Quil A (Spikoside; Iscotec AB)⁵. When spleen cells from BALB/c mice immunized subcutaneously in the base of the tail with ISCOMs containing 1 µg of recombinant HIV-1-IIIB gp160 envelope protein (ISCOM-gp160IIIB) were restimulated *in vitro* with mitomycin-C-treated syngeneic BALB/c fibroblasts transfected with the gp160IIIB gene, we obtained CTL that specifically killed fibroblast targets either expressing whole gp160IIIB envelope protein or pulsed with a 15-residue synthetic peptide (18IIIB) corresponding to an immunodominant CTL epitope (RIQRGPGRAFVTIGK) of gp160 (Fig. 1a)⁶. In contrast, similarly restimulated spleen cells from mice immunized with ISCOMs containing influenza proteins (Fig. 1b), or immunized with 1 µg of recombinant gp160IIIB mixed with complete Freund's adjuvant (Fig. 1d), incomplete Freund's adjuvant (Fig. 1e), or phosphate-buffered saline (Fig. 1f), or from unimmunized mice (Fig. 1c) did not demonstrate CTL activity against any of these targets. It seems, therefore, that exogenous soluble protein in other forms does not prime specific CTL. We could not detect any CTL activity against targets that were sensitized with a control peptide (T1)⁷ corresponding to a helper T-cell site of gp160 (Fig. 1a). The priming by ISCOM-gp160IIIB immunization lasts at least 4 months and has been consistently successful (in more than 10 experiments).

Because immunization with exogenous protein usually induces MHC class II-restricted CD4⁺ T cells, it was important

to determine the phenotype of the specific CTL primed by the ISCOM-gp160IIIB material and to map the MHC restriction of these effector cells. As Fig. 2 shows, the CTL were conventional CD8⁺CD4⁻ CTL (Fig. 2a) that recognized the immunodominant peptide 18IIIB determinant in association with the D^d MHC class I molecule (Fig. 2b). Similarly, antigen-specific CD8⁺CD4⁻ CTL that lysed MHC class II-negative fibroblast targets were also elicited after immunization with ISCOMs containing influenza haemagglutinin and restimulation with the influenza virus (Fig. 2c). As controls, ISCOM-gp160 did not induce influenza-specific CTL after influenza restimulation (Fig. 2c), and the haemagglutinin preparation itself only elicited marginal lytic activity, at least an order of magnitude less than that elicited by ISCOM immunization (data not shown). This is consistent with the recent report⁸ that two intranasal immunizations with ISCOMs containing the influenza A virus glycoprotein elicited a local increase in cytotoxic activity in the lung, although the CTL phenotype or MHC restriction was not tested in this case. Thus, this use of ISCOMs will be applicable to other proteins.

Although the effector cells resulting from ISCOM-antigen immunization and *in vitro* restimulation were CD8⁺CD4⁻, it was nevertheless possible that the effect of ISCOM-antigen immunization was merely to produce so much antigen-specific CD4⁺ T-cell help that we were observing the *in vitro* primary responses of CD8⁺ pre-CTL. We thought that this was unlikely because we have never observed *in vitro* sensitization of CD8⁺ CTL by stimulation of non-immune spleen cells with the gp160 transfectant even in the presence of exogenous interleukin-2 (data not shown), and because no such effect was seen following gp160 immunization in complete Freund's adjuvant (Fig. 1d), a potent inducer of CD4⁺ T-cell help. To test the possibility explicitly, we separated CD4⁺ and CD8⁺ T cells from spleen cells of unimmunized mice and mice immunized with ISCOM-gp160IIIB, remixed these in all four possible combinations, restimulated them with the gp160 transfectant, and assayed CTL activity on fibroblast targets transfected with the gp160 gene (Fig. 3), as well as on peptide 18IIIB-pulsed targets and control unpulsed targets (data not shown). The killing of peptide-pulsed targets exactly paralleled that on the gp160 transfectant, whereas no specific lysis (<3%) was observed on the control targets. The results showed that the CD8⁺ population must be primed *in vivo* by the ISCOM-gp160IIIB immunization to yield active CTL, and that the CD4⁺ population from the immunized mice is not capable of inducing CD8⁺ CTL from unprimed CD8⁺ cells during stimulation *in vitro* with the transfectant. In addition, because the combination of primed CD4⁺ cells with primed CD8⁺ cells induces the highest killing rates, the ISCOM immunization probably primes helper T cells as well as CTL.

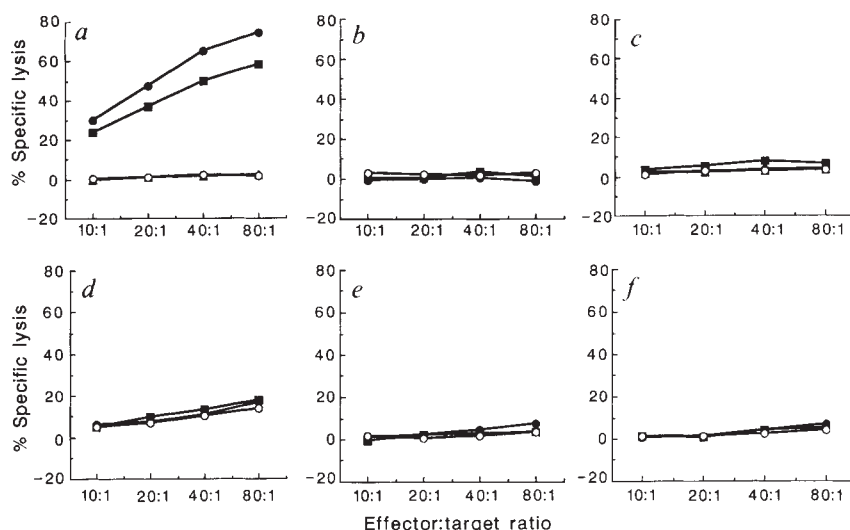
As it has been clearly shown by several groups that peptides can sensitize targets for killing by CTL^{6,9-11} and as the induction of ovalbumin-specific CTL by immunization with a tryptic digest of the protein has been reported¹², it was possible that peptide degradation products contaminated the gp160IIIB antigen preparation and thus might account for the CTL priming observed. The preparation was pure as judged by gels and western blots (see Fig. 1 legend), but these methods might not have detected small peptides. As a more relevant functional assay for such degradation products, we tested the same preparation of recombinant gp160IIIB used for ISCOM formation for its ability to sensitize cells to become CTL targets. Whereas 0.001 µM peptide was sufficient to sensitize targets for killing at near-maximal levels by a gp160IIIB-specific CTL line, no significant killing could be induced by 0.2 µM recombinant gp160IIIB protein (Fig. 4). Therefore, 0.5% contamination with peptide would be expected to produce maximal killing capacity, so that any functional contamination with peptide could not be more than one-tenth or one-hundredth of that, and was probably insufficient to account for the immunization observed. In addition, the same

[¶] To whom correspondence should be addressed at Building 10, Room 6B-12, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA.

[†] Present address: Department of Microbiology and Immunology, Nippon Medical School, Tokyo 113, Japan.

FIG. 1 Induction of HIV-1 envelope-specific CTL activity by immunization with recombinant gp160 in ISCOMs but not in conventional adjuvants. The fibroblast targets were either BALB/c.3T3 transfectants expressing the whole gp160IIIB envelope protein (black squares) or non-gp160-expressing BALB/c.3T3 cells pulsed with 1 μ M peptide 18IIIB (black circles) or with control peptide (T1) from a different region of the envelope protein⁷ (black triangles) or unpulsed (white circles).

METHODS. Spleen cells (5×10^6 ml⁻¹) (a) from BALB/c (H-2^d) mice subcutaneously immunized with 1 μ g of ISCOM-gp160IIIB in the base of the tail, (b) from mice immunized with ISCOMs containing influenza haemagglutinin, (c) from unprimed BALB/c mice, (d) from mice immunized with 1 μ g of recombinant gp160IIIB in complete Freund's adjuvant (DIFCO laboratories, Detroit, Michigan), (e) incomplete Freund's adjuvant, or (f) phosphate-buffered saline, were restimulated *in vitro* with 2×10^5 ml⁻¹ mitomycin-C-treated (50 μ g ml⁻¹, 45 min, 37 °C) gp160IIIB gene-transfected MHC-identical BALB/c.3T3 fibroblasts in 24-well culture plates in complete T-cell medium (a 1:1 mixture of RPMI 1640 and EHAA medium containing 10% fetal calf serum, 2 mM L-glutamine, 100 U ml⁻¹ penicillin, 100 μ g ml⁻¹ streptomycin and 5×10^{-5} M 2-mercaptoethanol). After culture for 6 days, cytolytic activity of the restimulated cells was measured using a 6 h assay⁶ with various ⁵¹Cr-labelled targets at the indicated effector:target ratios. The percentage specific release of ⁵¹Cr was calculated as $100 \times (\text{experimental release} - \text{spontaneous release}) / (\text{maximum release} - \text{spontaneous release})$. Maximum release was determined from supernatants of cells that were lysed by addition of 5% Triton-X 100. Spontaneous release was determined from target cells incubated without added effector cells. Standard errors of the means of triplicate cultures were always less than 5% of the mean. ISCOM-gp160IIIB and ISCOM-haemagglutinin were prepared using dialysis⁵. One mg of gp160IIIB or of influenza haemagglutinin at a concentration of 100 μ g ml⁻¹ was solubilized with 2% MEGA-10¹⁶, 100 μ g cholesterol, 100 μ g phosphatidyl choline and 500 μ g Quil A. After treatment in a sonicator bath for 20 min,



the mixture was left at room temperature for 2 h before overnight dialysis against phosphate-buffered saline at room temperature, and then at 4 °C for an additional 48 h. The mixture was then sedimented through 10% sucrose in a Kontron STS 28.38 rotor at 25,000 r.p.m. for 25 h at 15 °C. The pellet was dissolved in 1 ml phosphate-buffered saline. The ISCOMs were characterized as described previously⁵ by sedimentation in sucrose (19S), electron microscopy and sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Recombinant gp160 was prepared from cells infected with a recombinant baculovirus expressing the gene for gp160 of HIV-1-IIIB as described previously¹⁷. It was purified by lentil lectin chromatography and gel filtration¹⁷. The purified product was a single band on SDS-PAGE seen by both Coomassie blue staining and by western blotting with a monoclonal anti-gp41 antibody, although very faint bands at 110,000 and at about 60,000 were detected on an overexposed blot with HIV⁺ human sera.

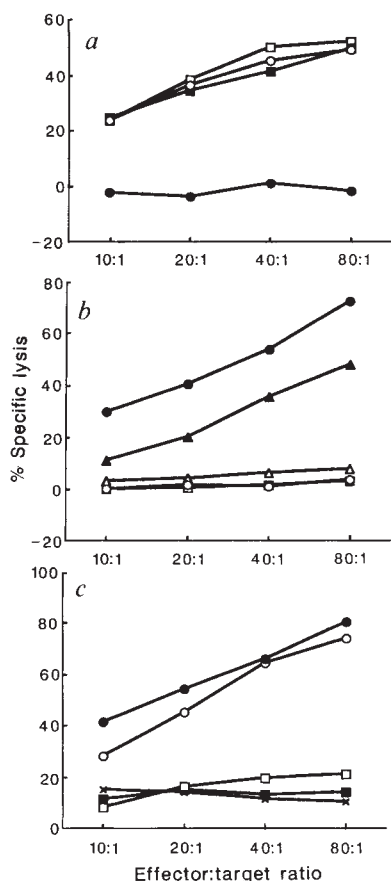


FIG. 2 Phenotype and MHC class I-restriction of the CTL induced by ISCOM-gp160IIIB or ISCOM-haemagglutinin immunization. a, The CTL induced by ISCOM-gp160IIIB immunization and by restimulation with gp160IIIB gene-transfected MHC-identical BALB/c.3T3 fibroblasts were treated with either anti-CD8 monoclonal antibody (3.155; rat IgM)¹⁸ plus complement (black circles) or anti-CD4 monoclonal antibody (RL.174; rat IgM)¹⁹ plus complement (black squares) or complement alone (white circles), or untreated (white squares), and tested on peptide 18IIIB-pulsed BALB/c.3T3 fibroblasts as targets. b, To determine the MHC restriction of the CTL induced by ISCOM-gp160IIIB, peptide 18IIIB-pulsed MHC class I-transfected L-cell fibroblasts expressing D^d (T4.8.3; black triangles)²⁰, L^d (T1.1.1; white triangles)²⁰, K^d (B4II2, L-K^d; white circles)²¹, and untransfected L-cells (white squares) were used as targets. As a positive control, BALB/c.3T3 fibroblasts (black circles) expressing all three H-2^d MHC class I molecules were used. (c) To test the generalizability of these results for another antigen, BALB/c mice were immunized as in Fig. 1 with 2 μ g ISCOM-influenza haemagglutinin, and 2 weeks later their spleen cells were restimulated for 6 days *in vitro* with syngeneic spleen cells infected with influenza A/PR/8/34(H1N1) virus (2000 HAU per 10^7 cells)²². The resulting effector cells were tested for lysis of MHC class II molecule-negative BALB/c.3T3 fibroblast targets infected with the H1-VAC vaccinia recombinant virus²² expressing the gene for the haemagglutinin of influenza A/PR/8/34 (H1N1) (10 plaque-forming units per cell) either after no treatment (white circles) or after treatment of effectors with anti-CD8 antibody (3.155; rat IgM)¹⁸ plus complement (black squares) or with complement alone (black circles). Alternatively, the ISCOM-haemagglutinin immune spleen cells were restimulated with gp160IIIB gene-transfected BALB/c.3T3 fibroblasts as a control (crosses). As a further control for the requirement for ISCOM-haemagglutinin immunization, spleen cells from mice immunized with ISCOM-gp160IIIB were restimulated with influenza A/PR/8/34-infected syngeneic spleen cells and tested on the same H1-VAC-infected targets (white squares). None of these effectors showed significant killing of BALB/c.3T3 fibroblast targets infected with a control vaccinia virus (data not shown).

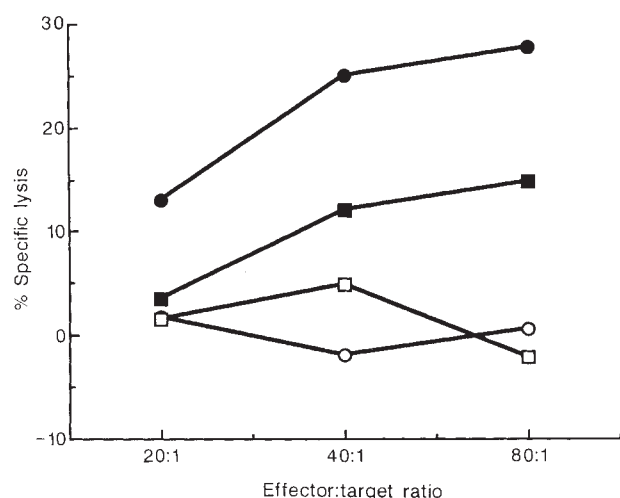


FIG. 4 To test for possible contamination of purified recombinant gp160IIIB protein with functional peptide degradation products that could sensitize targets for CTL, target BALB/c.3T3 fibroblasts were incubated overnight with several concentrations of the protein (0.2 μ M, white circles; 0.06 μ M, white squares; 0.02 μ M, white triangles). As positive controls, target fibroblasts were pulsed with various concentrations of peptide 18IIIB (0.1 μ M, black circles; 0.01 μ M, black squares; 0.001 μ M black triangles).

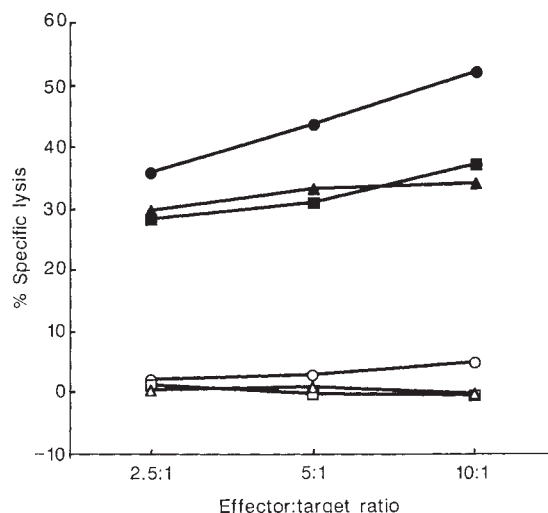
preparation of gp160 without ISCOMs did not elicit CTL (Fig. 1).

Thus, in contrast to results with other forms of exogenous soluble protein antigen, we have shown here that it is possible to prime antigen-specific, MHC-class I restricted CD8⁺ CD4⁻ CTL by immunization with exogenous intact protein using ISCOMs. Recent studies have suggested that MHC class I molecules predominantly present antigens that are synthesized in, or introduced into, the cytoplasm of the target cell, and that intact proteins taken up by the endosomal route are not normally able to give rise to MHC class I molecule-peptide complexes 'seen' by CD8⁺ T cells^{1,2,11}. Although ISCOMs do not fuse with cell membranes as liposomes can, they may be able to penetrate them and introduce antigen into the cytoplasm or even the endoplasmic reticulum, from which it can enter the MHC class I presentation pathway. Elucidating the mechanism by which CTL are induced by ISCOMs may shed new light on antigen-processing pathways for presentation by MHC class I molecules.

From this point of view, it is interesting that the CTL induced by immunization with ISCOM-gp160 kill targets that express the gp160 protein endogenously, with similar specificity to that induced by infection with a live vaccinia recombinant expressing gp160. CTL elicited by both types of immunization are dominated by cells specific for the peptide 18 antigenic site and CTL elicited by ISCOM-gp160IIIB, like those elicited by the recombinant gp160IIIB vaccinia virus^{13,14}, do not cross-react with targets expressing the gp160 gene of the variant isolates HIV-1 MN and HIV-1 RF (data not shown).

Recent data indicate that five humans immunized with the recombinant vaccinia virus expressing the HIV-1-IIIB envelope gp160 gene have specific CTL activity against the 18IIIB anti-

FIG. 3 Evidence for priming of CD8⁺ cells *in vivo* by ISCOM immunization. ISCOM-gp160IIIB-primed spleen cells and unprimed spleen cells were separated into CD4⁺ and CD8⁺ cells by treatment with either anti-CD8 monoclonal antibody (3.155)¹⁸ plus complement to eliminate CD8⁺ cells or anti-CD4 monoclonal antibody (RL172.4)¹⁹ plus complement to eliminate CD4⁺ cells. Mixtures of equal numbers (2.5×10^6 ml⁻¹) of ISCOM-primed CD4⁺ cells and primed CD8⁺ cells (black circles), primed CD4⁺ cells and unprimed CD8⁺ cells (white circles), unprimed CD4⁺ cells and primed CD8⁺ cells (black squares), or unprimed CD4⁺ cells and unprimed CD8⁺ cells (white squares) were cultured in 24-well plates with 2×10^5 ml⁻¹ of mitomycin-C treated gp160IIIB transfectants as stimulators for 5 days. Cytotoxicity was measured on ⁵¹Cr-labelled gp160IIIB transfectants (shown) and on untransfected BALB/c.3T3 fibroblasts pulsed or unpulsed with peptide 18IIIB (1 μ M). Results with peptide-pulsed targets exactly paralleled those with transfectants, whereas killing on unpulsed targets was consistently <3%. The s.e.m. of triplicate cultures was always <5% of the mean.



genic site¹⁵ originally identified in mice⁶, and that HIV-infected humans have MHC class I restricted CTL that kill targets pulsed with the peptide 18IIIB (M. Clerici, J. A. Berzofsky and G. M. Shearer, unpublished observations). Therefore, the experiments described here suggest that it may be possible to elicit human CTL by using ISCOMs containing HIV proteins. From a practical point of view, ISCOM-based vaccines may achieve the long-sought goal of induction of both CTL and antibodies by a purified protein. □

Received 16 October 1989; accepted 9 February 1990.

- Morrison, L. A., Lukacher, A. E., Braciale, V. L., Fan, D. P. & Braciale, T. J. *J. exp. Med.* **163**, 903-921 (1986).
- Germain, R. N. *Nature* **322**, 687-689 (1986).
- Staerz, U. D., Karasuyama, H. & Garner, A. M. *Nature* **329**, 449-451 (1987).
- Wraith, D. C. & Askonas, B. A. *J. gen. Virol.* **66**, 1327-1331 (1985).
- Morein, B., Sundquist, B., Höglund, S., Dalsgaard, K. & Osterhaus, A. *Nature* **308**, 457-460 (1984).
- Takahashi, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 3105-3109 (1988).
- Cease, K. B. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 4249-4253 (1987).
- Jones, P. D., Tha Hla, R., Morein, B., Lövgren, K. & Ada, G. L. *Scand. J. Immun.* **27**, 645-652 (1988).
- Townsend, A. R. M. *et al. Cell* **44**, 959-968 (1986).
- Maryanski, J. L., Pala, P., Corradin, G., Jordan, B. R. & Cerottini, J.-C. *Nature* **324**, 578-579 (1986).
- Moore, M. W., Carbone, F. R. & Bevan, M. J. *Cell* **54**, 777-785 (1988).
- Ishioaka, G. Y., Colon, S., Miles, C., Grey, H. M. & Chesnut, R. W. *J. Immun.* **143**, 1094-1100 (1989).
- Takahashi, H. *et al. Science* **246**, 118-121 (1989).
- Takahashi, H. *et al. J. exp. Med.* **170**, 2023-2035 (1989).
- Achour, A. *et al. Vth Int. Conf. on AIDS, Montreal* 546 abstract, (1989).
- Hildreth, J. E. K. *Biochem. J.* **207**, 363-366 (1982).
- Javaherian, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 6768-6772 (1989).
- Sarmiento, M., Glasebrook, A. L. & Fitch, F. W. *J. Immun.* **125**, 2665-2672 (1980).
- Ceredig, R., Lowenthal, J. W., Nabholz, M. & MacDonald, H. R. *Nature* **314**, 98-100 (1985).
- Margulies, D. H. *et al. J. Immun.* **130**, 463 (1983).
- Abastado, J.-P. *et al. J. exp. Med.* **166**, 327-340 (1987).
- Bennink, J. R. & Yewdell, J. W. *J. exp. Med.* **168**, 1935-1939 (1988).

ACKNOWLEDGEMENTS. We thank Dr Jack Bennink for gifts of H1-VAC recombinant vaccinia virus and influenza A/PR/8/34 (H1N1) virus, and Drs David Margulies and Gillis Otten for critical reading of the manuscript.

Novel cysteine-rich motif and homeodomain in the product of the *Caenorhabditis elegans* cell lineage gene *lin-11*

Gwen Frey^{d*}, Stuart K. Kim^{*†} & H. Robert Horvitz^{*}

^{*}Howard Hughes Medical Institute, Department of Biology, Room 56-629, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

[†]Department of Developmental Biology, Stanford Medical School, Stanford University, Stanford, California 94305, USA

THE gene *lin-11* is required for the asymmetric division of a vulval precursor cell type in the nematode *Caenorhabditis elegans*. Putative *lin-11* complementary DNAs were sequenced and found to encode a protein that contains both a homeodomain and two tandem copies of a novel cysteine-rich motif: C-X₂-C-X₁₇₋₁₉-H-X₂-C-X₂-C-X₂-C-X₇₋₁₁-(C)-X₈-C. Two tandem copies of this motif are also present amino-terminal to the homeodomains in the proteins encoded by the genes *mec-3*, which is required for *C. elegans* touch neuron differentiation¹, and *isl-1*, which encodes a rat insulin I gene enhancer-binding protein². The arrangement of cysteine residues in this motif, referred to as LIM (for *lin-11* *isl-1* *mec-3*), suggests that this region is a metal-binding domain. The presence in these three proteins of both a potential metal-binding domain and a homeodomain distinguishes them from previously characterized proteins.

During development, many cells must divide asymmetrically to produce daughters that differ in their fates. The nematode *C. elegans* permits analysis of asymmetric cell division, and the control of cell division has been extensively analysed at both the cellular and genetic levels³⁻⁷. The gene *lin-11* functions in the asymmetric division of the 2° vulval blast cells, which gener-

ate part of the hermaphrodite vulva^{4,5}. Whereas in wild-type animals each of the two 2° cells divides to produce daughter cells that express different fates, in *lin-11* mutant animals these cells produce daughters that express the same fate⁵ (Fig. 1a). Animals mutant for *lin-11* are unable to form a functional vulva and therefore cannot lay eggs⁴. In addition, *lin-11* males have abnormalities in tail morphology and animals of both sexes are slightly uncoordinated⁴, suggesting that *lin-11* has additional functions in *C. elegans* development.

We cloned *lin-11* by identifying a transposon insertion in the gene. A mutant allele of *lin-11* was isolated in the strain TR679, which has a high rate of Tc1 transposon transposition^{8,9}. To eliminate unlinked transposon insertions, we backcrossed animals carrying this new allele 10 times to a strain that has the wild-type number and pattern of Tc1 copies. To identify a transposon tightly linked to *lin-11*, we used flanking genetic markers to obtain chromosomes that had undergone recombination in the *lin-11* region. By examining 27 recombinants within a 4.5 map unit interval between the flanking markers *fer-1* and *unc-75*, we mapped a Tc1 insertion to within approximately 0.1 map units of *lin-11* (see legend to Fig. 1).

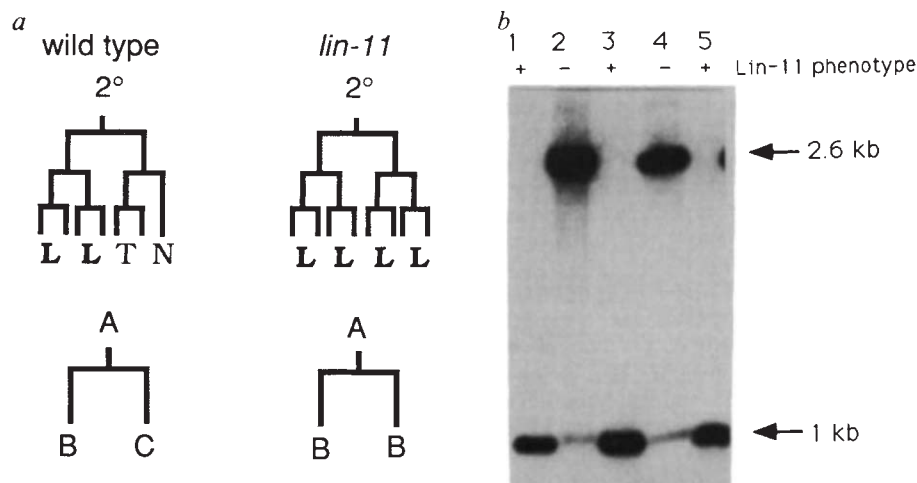
To establish that this transposon was responsible for the *Lin-11* phenotype, two strains were isolated in which the mutant phenotype had spontaneously reverted to wild type, and DNA from these strains was examined by Southern blot analysis. Both revertant strains lacked the transposon insertion (Fig. 1b), indicating that the transposon was inserted in a region of DNA required for *lin-11* function.

We cloned a 2.6-kilobase (kb) *Eco*RI fragment of genomic DNA containing the transposon insertion and isolated an 0.8-kb subclone of single-copy genomic sequence by removing the Tc1 and 0.2 kb of genomic DNA. We used this subclone to probe a cDNA library made from wild-type mixed-stage poly(A⁺)RNA (provided by J. Ahringer and J. Kimble) and obtained 16 cDNA clones of various sizes.

Strains bearing four of the five mutant *lin-11* alleles have DNA alterations in a region hybridizing to a putative *lin-11*

FIG. 1 a, Actual and schematic cell lineages of the wild-type and *lin-11* 2° vulval blast cell type (adapted from ref. 5). Horizontal lines indicate a cell division, and the vertical axis indicates time. N, no division; T, transverse division; L, longitudinal division and adherence of daughter cells to the ventral cuticle. A, B, C represent distinct cell types. b, Southern blot of wild-type, mutant, and revertant DNAs digested with *Eco*RI and probed with cloned single-copy genomic DNA that flanks the Tc1 transposon insertion. The wild-type restriction pattern is restored in spontaneous revertants, indicating that the loss of *lin-11* function in the mutant is caused by the transposon insertion. Lane 1, Bristol wild-type strain N2; lane 2, *lin-11*(*n1281::Tc1*) strain MT5195; lane 3, MT5204, wild-type phenotypic revertant of MT5195; lane 4, *lin-11*(*n1281::Tc1*) strain MT5201; lane 5, MT5202, wild-type phenotypic revertant of MT5201. The 1.0-kb *Eco*RI band is ~50-fold less intense than the 2.6-kb *Eco*RI band in the *Lin-11*(-) strains in lanes 2 and 4, indicating that the transposon has a low rate of somatic excision. +, Strain with wild-type *Lin-11* phenotype; -, strain with mutant *Lin-11* phenotype.

METHODS. b, DNA preparations²⁴ were subjected to Southern-blot analysis²⁵ using random-primed ³²P-labelled²⁶ plasmid. Spontaneous phenotypically non-*Lin-11* revertants were isolated from *lin-11*(*n1281::Tc1*) strains as hermaphrodites capable of laying eggs. These revertants (reversion rate <1 in 10⁵) were isolated during mapping of the Tc1 transposon insertion. *lin-11* is flanked by *fer-1* and *unc-29* ~1.5 map units to the left and by *unc-75* ~3 map units to the right. We scored 14 recombination events between *unc-75* and *lin-11*(*n1281::Tc1*) and 13 between *lin-11*(*n1281::Tc1*) and *unc-29*. 14/14 *Lin-11* and 0/13 non-*Lin-11* recombinant chromosomes contained the Tc1 insertion. Single-copy DNA flanking the *n1281::Tc1* transposon insertion was cloned by isolating DNA fragments between 1.5 and 3 kb in size from an agarose gel containing *Eco*RI-digested *n1281::Tc1*



genomic DNA, cloning the DNA into λ gt10 (ref. 25), and screening the library with pTc1 (ref. 27). A 2.6-kb Tc1-hybridizing insert was isolated, recombined into pBS⁺ (Stratagene) and designated pGF1. The 1.6-kb Tc1 insert and 0.2 kb of the flanking genomic DNA were deleted from pGF1 by digestion with *Eco*RV followed by religation, generating a plasmid, pGF2, that contains an insert of 0.8 kb of single-copy genomic DNA that had flanked the transposon insertion. The pGF2 0.8-kb *Eco*RI fragment was isolated from an agarose gel and used for labelling. Lane 2, *lin-11*(*n1281::Tc1*) strain isolated as a homozygous *fer-1*(+) *unc-75*(e950) recombinant segregant from an *unc-75* non-*fer-1* recombinant progeny of a *lin-11*(*n1281::Tc1*)/*fer-1*(*hc1ts*) *unc-75*(e950) parent; lane 4, *lin-11*(*n1281::Tc1*) strain isolated as a homozygous *unc-29*(e1072) *unc-75*(+) segregant from an *unc-29* non-*unc-75* recombinant progeny of a *lin-11*(*n1281::Tc1*)/*unc-29*(e1072) *unc-75*(e950) parent.

cDNA. We used a 770-base pair (bp) subclone of a 1.7-kb cDNA clone to probe genomic DNA from strains bearing *lin-11* alleles, and found that in addition to the transposon allele, three of four ethyl methanesulphonate-derived alleles display altered patterns of restriction fragments (Fig. 2a). DNA from the mutant strain *n389* does not hybridize to the putative *lin-11* cDNA probe and thus seems to be deleted for the sequence encoded by this cDNA subclone. The mutant strains *n566*, *n672* and *n1281::Tc1* are altered in three different, adjacent fragments that hybridize to this cDNA (Fig. 2b), strongly suggesting that this cDNA corresponds to *lin-11*.

We determined the sequence of this 1.7-kb cDNA, which contains an open reading frame encoding at least 382 amino acids (Fig. 3). A methionine at amino acid position 27 could define a translational initiation site. We have not established the 5' end of the transcription unit, however, and this cDNA might be incomplete for *lin-11* coding sequences. Preliminary northern blot analysis has identified a single *lin-11* transcript of about 1.7 kb (data not shown).

The amino-acid sequence of the *lin-11* open reading frame contains three regions of interest—a homeodomain, a potential transcriptional activation domain, and a domain containing two tandem copies of a novel cysteine-rich motif. The *lin-11* sequence has extensive similarity to the protein products of two other genes, each of which also contains three corresponding domains: *mec-3*, which is required for the differentiation of the mechanosensory neurons of *C. elegans*¹, and *isl-1*, which

encodes a protein that binds the rat insulin I gene enhancer region².

The *lin-11* homeodomain, which has eight of the nine most conserved amino acids characteristic of homeodomains¹⁰ (Fig. 4a), is most similar to the homeodomain encoded by *mec-3* and is also similar to that encoded by *isl-1*. Homeodomains are involved in DNA binding and have been found in proteins that regulate gene expression¹⁰. The homeodomain-encoding region of *lin-11* is interrupted by the insertion of the *Tc1* transposon in the *lin-11(n1281::Tc1)* allele (Fig. 3), and the site of the insertion directly precedes a region of the homeodomain implicated in the recognition of DNA¹¹.

Carboxy-terminal to the homeodomain in the deduced *lin-11* protein product is a region rich in proline residues. Carboxy-terminal to the homeodomain of the *mec-3* gene product is an acid-rich region¹, and carboxy-terminal to the homeodomain of *isl-1* is a glutamine-rich region². Proline-rich, acid-rich and glutamine-rich regions can all act in transcriptional activation when present in genes that regulate gene expression¹².

Amino-terminal to the homeodomain in the deduced *lin-11* protein product is a cysteine-rich region containing two tandem copies of the motif C-X₂-C-X₁₇₋₁₉-H-X₂-C-X₂-C-X₂-C-X₇₋₁₁-(C)-X₈-C (Fig. 4b). The *isl-1* and *mec-3* gene products also each have two tandem copies of this motif, which we call LIM, for *lin-11* *isl-1* *mec-3*. In the *lin-11*, *mec-3* and *isl-1* gene products. The tandem LIM motifs are all present amino-terminal to, but at a variable distance from, the homeodomain (62 amino acids in *lin-11*, 74 in *mec-3*, and 50 in *isl-1*). A third *C. elegans* gene, *ceh-14* (ref. 13), which encodes a protein with a homeodomain

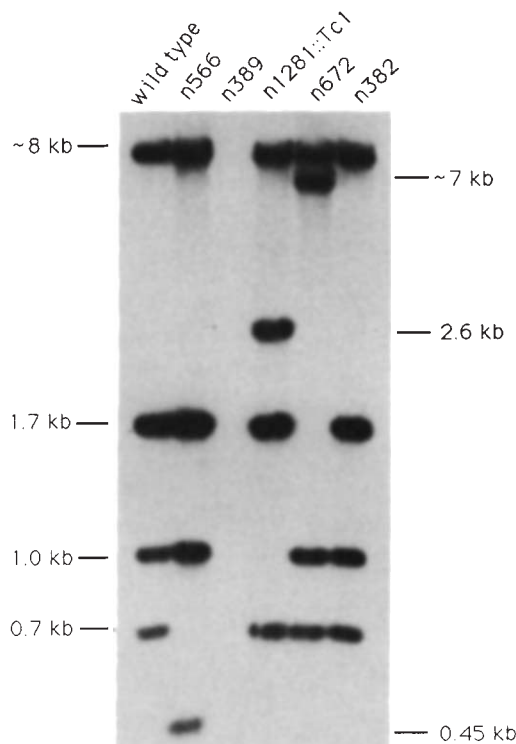
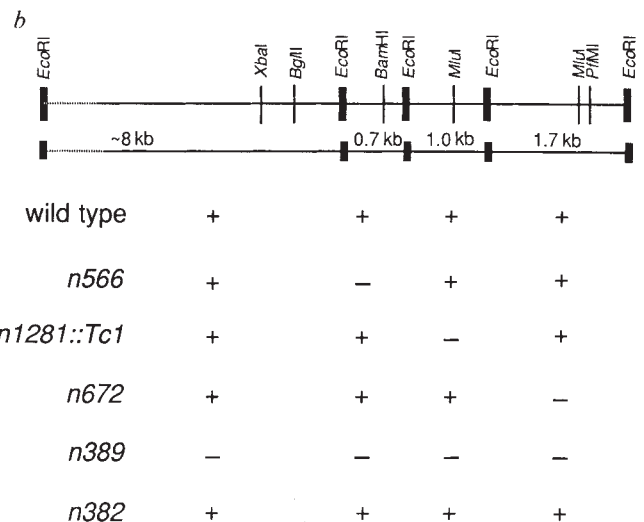


FIG. 2 a, DNA alterations associated with *lin-11* alleles. Southern blot of genomic DNA from strains bearing the five known mutant *lin-11* alleles⁴ digested with *EcoRI* and probed with a ³²P-labelled subclone of the cDNA, a partial sequence of which is shown in Fig. 3. Four of the five *lin-11* alleles, *n566*, *n389*, *n672* and *n1281::Tc1*, result in alterations when compared with the wild type. Approximate sizes on the left correspond to wild-type bands, and on the right to mutant-specific bands. All *lin-11* alleles except *n1281::Tc1* were isolated after mutagenesis with ethyl methanesulphonate⁴. b, Partial genomic map and diagram of location of *lin-11* allele-specific *EcoRI* restriction fragment patterns. The fragments of genomic DNA that hybridize to the cDNA subclone probe are indicated. The *lin-11* homeobox is contained in the 1.0- and 1.7-kb *EcoRI* genomic fragments and the LIM motifs in the 8-kb *EcoRI* fragment. +, Wild-type length restriction fragment present. -, Wild-type length fragment not present (hybridizing genomic sequences either rearranged or deleted).



METHODS. a, The cDNA from which the subclone probe was obtained, cGF43, was isolated by probing with pGF1 a λ gt10 cDNA library made from mixed stage poly(A⁺) RNA provided by J. Ahringer and J. Kimble. Clone cGF43 was partially digested with *EcoRI*, and DNA was purified from an agarose gel fragment and subcloned into pBS⁺ (Stratagene). This subclone, pGF4, which contains base pairs 1-1,082 as shown in Fig. 3, was further modified by digestion with *HincII* and religation of the remaining fragment and vector. The resulting clone, pGF5, contains the region between base pairs 306 and 1,082. A 1.1-kb *PvuII* fragment derived from pGF5 was used for labelling. As assayed by ethidium bromide staining, approximately equal amounts of DNA were loaded in each lane. Subclone pGF4 consists of exon sequences only and therefore deletions or insertions in the mutant alleles in non-coding sequence might not be detected by this assay. The probe does not contain the complete coding region of the cDNA, so we do not know if the *n389* mutant strain is deleted for the entire coding region of the gene. b, The genomic restriction map was constructed by analysis of the lengths of restriction fragments derived from cloned wild-type genomic DNA cut by the enzymes *EcoRI*, *MluI*, *XbaI*, *BamHI*, *BglII* and *PstI* in single and double digestions and by probing Southern blots containing *EcoRI*-digested wild-type DNA with the ³²P-labelled cDNA subclones pGF4 and pGF5.

FIG. 3 Partial nucleotide and deduced amino-acid sequence (single-letter code) of a *lin-11* cDNA. Shown below is a schematic diagram of features of the structure of the deduced *lin-11* protein product. Only sequences found in at least one other independently isolated cDNA are shown. The homeodomain is boxed, the cysteine and histidine residues defining the LIM motifs are circled, the borders of the two LIM motifs are indicated with brackets, the proline residues in the C-terminal region are indicated by asterisks, and the Tc1 transposon insertion site after the 29th amino acid in the homeodomain is indicated by a stippled triangle.

METHODS. The DNA sequence of both strands of a 1.7-kb cDNA clone, cGF43, was determined by the dideoxy chain-termination technique, using Sequenase enzyme (US). Partial sequences of one strand of each of three other cDNA clones, cGF41, cGF44 and cGF77, were also determined. We found no evidence of alternative splicing, and all clones seemed to represent the same transcript. However, at least 14 of the 16 cDNAs are incomplete. The 1,277-bp sequence shown appeared both in cGF43 and in other independently isolated cDNA clones. The ~400 bp of sequence in cGF43 5' to the displayed sequence were not found in any of the other 15 independently isolated cDNAs. The transposon insertion site was determined by sequencing ~100 bp of genomic DNA flanking the transposon in plasmid pGF1, which contains the 2.6-kb Tc1-hybridizing insert.

```

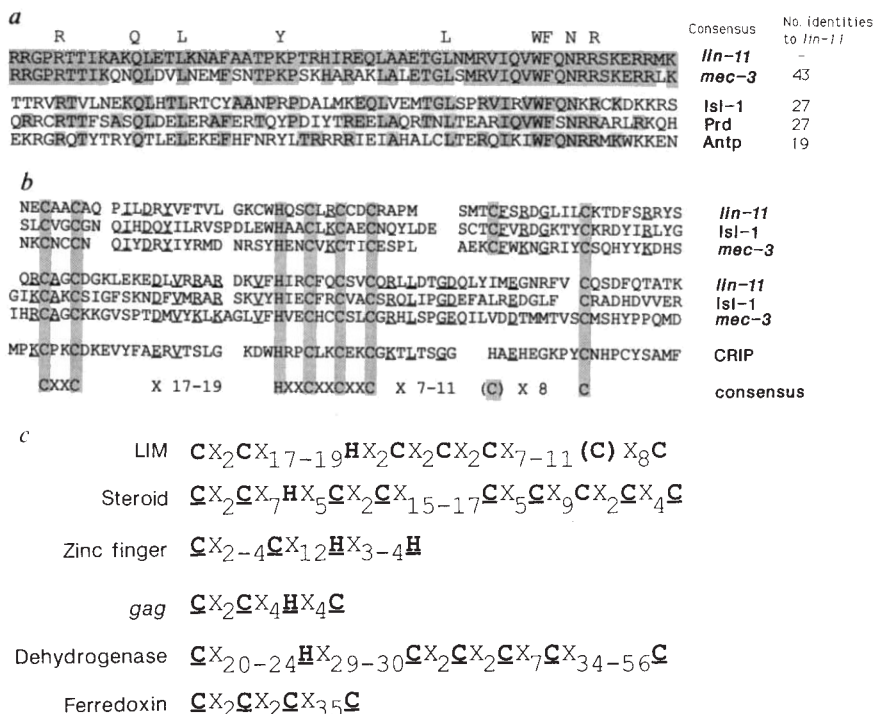
1  L H O Q Q S I E D V G S V T S S A T L L L L D S A T W M M P S
   CTCTTCATCAGCAGTCAATGAAGACGTGCGGAGTGCACCTCATCTGCCAGTTGCTGCTTCGACTCTGCCAGCTGGATGATGCCAT 90
91  S T T Q P H I S E I S G N E ( C A A C A Q P I L D R Y V F T V
   CGTCAACGACTCAGCCGACATCTCTGAATAGTGAATGATGTGCGCGGTGTCAGCAGCTATCTTCGACATGATGATTCATC 180
181 L G K C W ( H ) Q S C L R ( C D C R A P M S M T C F S R D G L I
   TGCTCGGAAATGCTGGCATCAGTCTGTCTGCTGATGTGCTGCTGCGAGCCCATGCTCAATGACTGTCTTATAGATGATGCTT 270
271 L ( C ) K T D F S R R Y S Q R R ( C A G C D ) G K L E K E D L V R R A
   TTCTGTGTAAACATGATTTTCAAGCGCATATAGTCAACGGTGTGCGGATGCGATGGAAATGGAAAAAGAGGATCTGGTGAGACG 360
361 R D K V F ( H ) I R C F Q C S V ( C ) Q R L L D T G D Q L Y I M E G
   CAAGAGATAAAGTTTTCATATTCGATGTTTTCATGCTCGGTTTGTCAAGGTTTATGACACATGCTGACGAGCTTTATATCATGGA 450
451 N R F V ( C ) Q S D F Q T A T K T S T P T S I H R P V S N G S E
   GCAATCGATTCTGTGTCAAAGTATTTTCAACAGCAACCAAAACATCAACGCCACATCGATTCATAGCCCGATTCATCAATGATCCG 540
541 C N S D V E E D N V D A C D E V L G D D D G E G D C G K D N S
   AATCAATTCGATTCGAGGAGGATGTGATGCTTGTGATGAGTGGATCGATGATGGAGAGGCGATGTTGGAGGACCAAT 630
631 D D S N S A K R R G P R T T I K A K Q L E T L K N A F A A T
   CAGATGACTCGAATCTGCGAAAGGCGCTCGACGACCAATTAAGCAAAACATCTGAACCTTAAAAACCGGCTTCGCGCCAA 720
721 P K P T R H I R E Q L A A E T G L N M R V I Q V W F N R R
   CACCAACCAACTCGACATTCCTGTGAACAACTTGCCTGCGGAGACAGGACTCAACATGAGTCTTCAGGTGTGGTTTCAAAACCG 810
811 S K E R R M X Q L R F G G Y R Q S R R P R R D I V D M F P
   GAAGCAAGAAACGAAGATGAACAACTTCGTTTGAAGGATATGCTCAATCCGAGACCAAGCCGCTGACGATATTGTAGATATG 900
901 N D Q Q F Y P P P P S N V Q F F C D P Y T T S P N P E T
   CGAATGACCAACATCTACCCACCACTCCACCATCAATGTTCAATCTTTTGTGACCATACACAACTTCCCAACATCCAGAAA 990
991 I Q M A Q P F A V T E N M N M V P E Y T E Q S A T P P E
   CATTCAATGCTCCCAATTCGATTCCTCAACGAAAATGAAACATGGTCCGAGGACCAATGAGCAATCAGGACCCCGCAG 1080
1081 F N E D T F A C I Y S T D L G K P T P V S W S t o p
   AATTCAATGAAGACATTTGATGATGATTTTCAACAGACTTGGGAAACCACTCCGGTTTCATGTTAGCAGCTCGACCTCCCAACA 1170
1171 ATCCACATACAGGAGTATTCACGTTCTGTAGTGTCTTCTCTCGCATATTTTAAATCTCAAAAGCACTCACTTAAACATTTTCCA 1260
1261 CACCGATTTCCAATGTC 1277

```



FIG. 4 *a*, Sequence comparison of homeodomains (single-letter amino-acid code). The *lin-11* homeodomain is aligned with the homeodomains of *mec-3* (ref. 1) and *isl-1* (ref. 2) as well as with those of the *Drosophila* proteins Paired (Prd)²⁸ and Antennapedia (Antp)²⁹. Shading indicates amino-acid identity with Lin-11. The nine most conserved *Drosophila* homeodomain amino acids¹⁰ are indicated at the top. The LIM homeodomains have a basic amino acid at residue 25 instead of the otherwise highly conserved tyrosine. If the numbers of conserved basic or acidic residues are added to the numbers of identities, the homeodomain of Lin-11 is more similar to Mec-3 (45/60) and Isl-1 (34/60) than to any other homeodomains (for example 30/60 for Prd). None of the LIM protein homeodomains falls into any of the previously identified classes of homeodomains^{1,2,10}. Lin-11 shares similarities with *prd* gene products, but does not contain a serine at position 50 or a number of other residues that are highly conserved in the *prd* class^{10,28}. *b*, Sequence comparison of the LIM motifs of Lin-11, Mec-3, Isl-1 and CRIP. The cysteine and histidine residues conserved among the four proteins are shaded. Conserved residues (identical or similarly charged) in the first (top three lines) or second (next three lines) LIM repeats of all three homeodomain-containing genes *lin-11*, *mec-3* and *isl-1* or among these proteins and CRIP in the second LIM repeat are underlined. In all four proteins the residue preceding the conserved histidine contains a phenyl group. CRIP has only one copy of the LIM motif, whereas the other three proteins have two copies. All amino acids between the first and second LIM repeats in Lin-11, Mec-3 and Isl-1 are shown. *c*, Highly conserved cysteine and histidine residues in the LIM motif and in five consensus motifs from metal-binding proteins. Underlined residues have been shown to coordinate binding of a metal ion in at least some members of the family from which the consensus is derived. The number of amino acids separating the cysteine and histidine residues is indicated. X, amino acid residue; Steroid, steroid hormone receptor proteins²²; Zinc finger, TFIIIA-like zinc-finger proteins^{19,20}; *gag*, retroviral *gag* gene-encoded core nucleic acid-binding proteins³⁰; dehydrogenase, dehydrogenase proteins³¹; ferredoxin, bacterial ferredoxin proteins³².

METHODS. *b*, A search of the National Biomedical Research Foundation (NBRF) Protein Sequence Database (release 22.0) and NBRF New Protein Sequence Database (release 40.0) using the program FASTP with the Lin-11



protein sequence identified only CRIP and Mec-3 as having significant identity to Lin-11. A search of these databases as well as the Owl 6.3 protein database for the partial LIM sequence H-X₂-C-X₂-C-X₂-C using the search program Match or FSEQ did not identify any additional proteins containing the complete LIM motif, although several other proteins do contain the sequence H-X₂-C-X₂-C-X₂-C or slight variants (such as H-X₂-C-X₂-C-X₂-C-X₃-C). The first 16 amino acids of the first LIM repeat in Mec-3 sequence shown differs from the published amino-acid sequence¹. The sequence presented here was determined from our identification of other potential splice patterns based on consensus splice sites used in *C. elegans*³³ and the published nucleic-acid sequence of *mec-3* genomic DNA¹. A possible exon in the *mec-3* genomic sequence from base pairs 2,493 to 2,544 would produce the sequence shown if spliced to an exon beginning at base pair 2,686 (ref. 1). Subsequent analysis has confirmed the existence of this splicing pattern in 6/6 *mec-3* cDNAs analysed (D. Xue and M. Chalfie, personal communication).

very similar to those of *lin-11* and *mec-3*, also has two copies of the LIM motif (T. Bürglin and G. Ruvkun, personal communication).

The LIM motif is also present in the deduced protein product of the cysteine-rich intestinal protein (CRIP) gene, which was isolated in a screen for RNAs of the small intestine that change concentration in the transition from suckling to weaning in rats¹⁴. In contrast to the other LIM proteins, the 77-amino-acid CRIP protein contains only a single copy of the LIM motif and lacks a homeodomain.

In addition to the cysteines, several other residues are conserved among the *lin-11*, *mec-3* and *isl-1* gene products in the region containing the LIM motifs (Fig. 4b). The pattern of conserved residues suggests that these amino acids are functionally important and that the first LIM domain is distinct from the second. The level of amino-acid identity within the region of 111 amino acids between the first cysteine residue of the first LIM motif and the last cysteine residue of the second LIM motif is higher between the *C. elegans* gene *lin-11* and the mammalian gene *isl-1* (48 amino acids identical) than between the two nematode genes *lin-11* and *mec-3* (35 amino acids identical).

Other conserved regions associated with homeodomains include the paired box¹⁵, a region of similarity in proteins that contain engrailed-class homeodomains¹⁶, and the POU-specific domain¹⁷. Like the paired box¹⁸, the LIM motif is also found in a protein that does not contain a homeodomain. Like the POU protein homeodomains, the LIM protein homeodomains do not readily fall into any previously defined class of homeodomain¹⁰. In addition, the LIM protein homeodomains share several features, including the presence of a basic amino acid instead of the otherwise highly conserved tyrosine residue at position 25 in the homeodomain (Fig. 4a).

Conserved cysteine residues separated by two amino acids are found in the metal-binding regions of a variety of proteins^{19,20} (Fig. 4c). On this basis, we suggest that the LIM motif functions in metal binding. (J. Way and M. Chalfie, personal communication, noted previously that the cysteine-rich region of *mec-3* may function in metal binding.) However, the LIM motif is not identifiable as any previously known metal-binding consensus motif, such as those found in the TFIIIA-class of zinc fingers²¹ or in the steroid hormone receptor family²².

Metal ions can be used by proteins to stabilize structures involved in protein-DNA, protein-RNA, or protein-protein interactions^{19,20}. If the LIM region were involved in a protein-DNA interaction, it might increase the specificity of the DNA binding conferred by the homeodomain or it might bind a second site on the DNA. If the LIM region were involved in a protein-protein interaction, it might contact other components of the transcriptional machinery or chromosomal or cytoskeletal proteins, or be involved in protein dimerization.

What does the structure of the *lin-11* protein product indicate about its role in asymmetric cell division? A second LIM gene, *mec-3*, has been postulated to function in an asymmetric cell division¹. (The *mec-3* gene also functions in at least one cell division in which it is not required to make sister cells different from each other²³.) The presence of the homeodomain and a potential transcriptional activation domain suggests that the products of *lin-11* and *mec-3* bind DNA and regulate the expression of genes specific to one of the two daughter cells generated during asymmetric cell division. Perhaps asymmetric cell divisions involving *lin-11* and *mec-3* require the production of the protein products of these genes in the mother cell and the subsequent segregation of these proteins to only one of the two daughter cells. If so, a possible function of the LIM region might be to bind to a protein or to a nucleic acid that is differentially segregated during cell division, resulting in the asymmetric segregation of a cell type-specific transcription factor. □

Received 28 December 1989; accepted 12 March 1990.

- Way, J. C. & Chalfie, M. *Cell* **54**, 5–16 (1988).
- Karlsson, O., Thor, S., Norberg, T., Ohlsson, H. & Edlund, T. *Nature* **344**, 879–882 (1990).
- Wood, W. B. *The Nematode Caenorhabditis Elegans* (Cold Spring Harbor Laboratory, New York, 1988).
- Ferguson, E. L. & Horvitz, H. R. *Genetics* **110**, 17–72 (1985).
- Ferguson, E. L., Sternberg, P. W. & Horvitz, H. R. *Nature* **326**, 259–267 (1987).
- Sternberg, P. W. & Horvitz, H. R. *Cell* **44**, 761–772 (1986).
- Sternberg, P. W. & Horvitz, H. R. *Dev. Biol.* **130**, 67–73 (1988).
- Kim, S. K. & Horvitz, H. R. *Genes Dev.* **4**, 357–371 (1990).
- Collins, J., Saari, B. & Anderson, P. *Nature* **328**, 726–728 (1987).
- Scott, M. P., Tamkun, J. W. & Hartzell III, G. W. *BBA Reviews on Cancer* **9**, 25–48 (1989).
- Gehring, W. J. *Science* **236**, 1245–1252 (1987).
- Mitchell, P. J. & Tjian, R. *Science* **245**, 371–378 (1989).
- Bürglin, T. R., Finney, M., Coulson, A. & Ruvkun, G. *Nature* **341**, 239–243 (1989).
- Birkenmeier, E. H. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2516–2520 (1986).
- Bopp, D., Burri, M., Baumgartner, S., Frigerio, G. & Noll, M. *Cell* **47**, 1033–1040 (1986).
- Joyner, A. L., Kornberg, T., Coleman, K. G., Cox, D. R. & Martin, G. R. *Cell* **43**, 29–37 (1985).
- Herr, W. *et al. Genes Dev.* **2**, 1513–1516 (1988).
- Bopp, D., Jamet, E., Baumgartner, S., Burri, M. & Noll, M. *EMBO J.* **8**, 3447–3457 (1989).
- Berg, J. M. *Science* **232**, 485–487 (1986).
- Berg, J. M. in *Cold Spring Harbor Symp. quant. Biol.* **52** (Cold Spring Harbor Laboratory, New York, 1987).
- Klug, A. & Rhodes, D. *Trends biochem. Sci.* **12**, 464–469 (1987).
- Evans, R. M. *Science* **240**, 889–895 (1988).
- Scott, M. P., Tamkun, J. W. & Hartzell III, G. W. *BBA Reviews on Cancer* **9**, 25–48 (1989).
- Way, J. C. & Chalfie, M. *Genes Dev.* **3**, 1823–1833 (1989).
- Sulston, J. & Hodgkin, J. in *The Nematode Caenorhabditis Elegans* (ed. Wood, W. B.) 587–606 (Cold Spring Harbor Laboratory, New York, 1988).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1989).
- Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6–13 (1983).
- Finney, M., Ruvkun, G. & Horvitz, H. R. *Cell* **55**, 757–769 (1988).
- Frigerio, G., Burri, M., Bopp, D., Baumgartner, S. & Noll, M. *Cell* **47**, 735–746 (1986).
- Schneuwly, S., Kuroiwa, A., Baumgartner, S. & Gehring, W. J. *EMBO J.* **5**, 733–739 (1986).
- Green, L. M. & Berg, J. M. *Proc. natn. Acad. Sci. U.S.A.* **86**, 4047–4051 (1989).
- Aronson, B. D., Somerville, R. L., Eppery, B. R. & Dekker, E. E. *J. biol. Chem.* **264**, 5226–5232 (1989).
- Bruschi, M. H., Guerlesquin, F. A., Bovier-Lapierre, G. E., Bonicel, J. J. & Couchoud, P. M. *J. biol. Chem.* **260**, 8292–8296 (1985).
- Emmons, S. in *The Nematode Caenorhabditis Elegans* (ed. Wood, W. B.) 47–79 (Cold Spring Harbor Laboratory, New York, 1988).

ACKNOWLEDGEMENTS. We thank T. Bürglin, E. Ferguson, A. Frankel, A. MacMurray, G. Ruvkun, D. Turner, and the members of the Horvitz laboratory for helpful discussions. We also thank J. Sulston and A. Coulson for generously providing cosmid clones, J. Ahninger and J. Kimble for providing a cDNA library, M. Chalfie for communicating unpublished results, and especially T. Bürglin for sharing his observation that a homeobox probe hybridized to *lin-11* cosmids, for generously providing the homeobox probe, and for communicating unpublished results. This work was supported by the United States Public Health Service. H.R.H. is an Investigator of the Howard Hughes Medical Institute. G.F. was an Office of Naval Research Graduate Fellow and is supported by the NIH. S.K. was a fellow of the Helen Hay Whitney Foundation and is a Lucille P. Markey Scholar in Biomedical Science.

Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain

Olof Karlsson, Stefan Thor, Torbjörn Norberg, Helena Ohlsson & Thomas Edlund*

Department of Microbiology, University of Umeå, S-901 87 Umeå, Sweden

THE activity of the rat insulin I gene enhancer is mainly dependent on two *cis*-acting protein-binding domains^{1–3}. Here we report the isolation of a complementary DNA encoding a protein, Isl-1, that binds to one of these domains. Isl-1 contains a homeodomain with greatest similarity to those of the *Caenorhabditis elegans* proteins encoded by *mec-3* (ref. 4) and *lin-11* (ref. 5). In addition, Isl-1, like the *lin-11* and *mec-3* gene products, contains a novel Cys-His domain which is reminiscent of known metal-binding regions. Together these proteins define a novel class of proteins containing both a homeo- and a Cys His-domain. Isl-1 is preferentially expressed in cells of pancreatic endocrine origin. If the structural homologies between Isl-1 and the *C. elegans* gene products reflect functional similarities, a role for Isl-1 in the development of pancreatic endocrine cells could be envisaged.

We screened a cDNA λgt11 library derived from poly(A)-containing RNA from the insulin-producing RIN cells^{6,7} using

* To whom correspondence should be addressed.

a probe consisting of four tandem copies of the insulin gene enhancer sequence of positions -199 to -252. From a primary screen of about 600,000 plaques, we isolated one recombinant clone that bound this probe and carried a 0.9-kilobase (kb) insert.

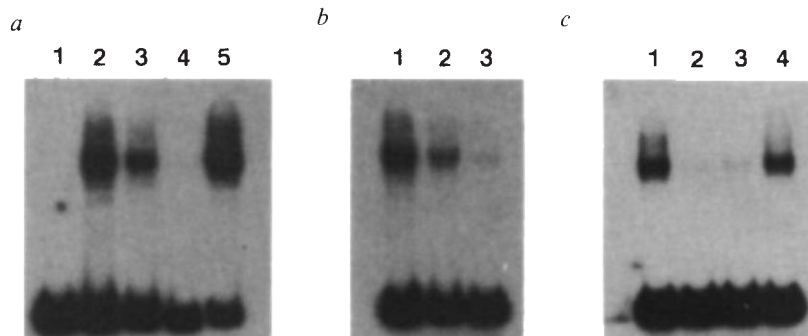
We expressed the coding portion of this insert recombinant clone in *Escherichia coli* as a fusion product with the TrpE protein^{8,9}. This fusion protein was shown by electrophoretic mobility shift assay (EMSA) to bind to a DNA fragment containing the -199 to -252 enhancer sequence (Fig. 1a). We identified the site of interaction in this sequence by analysing binding to

probes carrying the S19, S20 and S21 block mutations¹. Binding was unaffected with the S21 mutant, was partial with mutant S19, and absent with the S20 mutant. This shows that most of the nucleotides in the binding site are located in the region encompassing positions -211 to -222 of the insulin gene enhancer. In the -222 to -208 sequence, TTAATAA-TCTAATTA, there are three alternative copies of a TAAT or ATTA motif, which is part of binding sites for homeodomain-containing proteins^{10,11}. In mutant S21, none of the motifs are mutated, in S20 all three motifs are mutated, and in S19 one

FIG. 1 Specific binding of bacterially synthesized TrpE/IsI-1 fusion protein to insulin gene enhancer sequences. *a*, Binding of the fusion protein was analysed by EMSA using wild-type and mutant variants of the -199 to -252 enhancer sequence as radiolabelled probes. The block mutants S19, S20 and S21 are described in ref. 1 and cover sequences -199 to -210, -211 to -222 and -223 to -232, respectively. Lane 1, control extract and wild-type probe; lane 2, TrpE/IsI-1 extract and wild-type probe; lane 3, TrpE/IsI-1 extract and S19 probe; lane 4, TrpE/IsI-1 extract and S20 probe; lane 5, TrpE/IsI-1 and S21 probe. *b*, EMSA was performed as in *a* using TrpE/IsI-1 extract and wild-type rat insulin I gene probe (lane 1), sequences from the rat insulin II gene enhancer corresponding to the -199 to -252

sequences of the rat insulin I gene enhancer as a probe (lane 2), or rat insulin I gene probe containing three point mutations at positions -212, -217 and -219 (lane 3). *c*, Competition analysis. Binding of the TrpE/IsI-1 fusion protein to the wild-type rat insulin I gene probe was competed by a 500 M excess of unlabelled probe. Lane 1, no competitor; lane 2, wild-type rat insulin I gene probe; lane 3, wild-type rat insulin II gene probe; lane 4, S20 block mutant probe.

METHODS. A λ gt11 oligo(dT)-primed cDNA library prepared from RIN cells⁶ was screened with a DNA fragment containing four tandem copies of the -252 to -199 rat insulin I gene enhancer sequence essentially according



to Sing *et al.*⁷. In a primary screen of 600,000 clones, only one bound specifically to this fragment. The recombinant carried a 960-bp cDNA insert which was inserted into the TrpE expression plasmid path11, and the induction of fusion protein expression and the preparation of extract was essentially according to Klempner *et al.*⁸. In the EMSA, 0.3 μ g *E. coli* extract was incubated for 15 min at room temperature in a 10- μ l reaction mixture containing 150 mM KCl, 25 mM HEPES buffer, pH 7.9, 15% glycerol, 5 mM DTT, 3 μ g poly(dI-dC) and 1-3 fmol end-labelled DNA fragment and processed as previously described³. Competing DNA was added to the mix 15 min before adding the probe.

CAGGGGCCACTATTGGCCACCTAGCCACAGCAGCATCTCTGTGGGCTGTTACCAATTGT 66
M G D M G D P P
CCAACCACCATTTCACTGTGGACATTACTCCCTCTTACAGATATGGGAGACATGGCGATCCACCA 132
K K K R L I S L C V G C G N Q I H D Q Y I L
AAAAAAAGCTGCTGATTTCCCTATGTGTGGTGGGTAATCAAAATCAGCATCAGTATATTCTG 198
R V S P D L E W H A A C L K C A E C N Q Y L
AGGGTTCTCCGGATTGGAAATGCGATGCGGATGTTTGAATATGCGGAGTGATCAATCAGTATTG 264
D E S C T C F V R D G K T Y C K R D Y I R L
GACGAAAGCTGTACCTGCTTTGTTAGGACGGGAAACCTACTGTAAAGAGATATATCAGGTG 330
Y G I K C A K C S I G F S K N D F V M R A R
TACGGGATCAAAATGCGCAAGTGCAGCATAGGCTTCAGCAAGAAGCAGCTTCGTGATGCGCGCCG 396
S K V Y H I E C F R C V A C S R Q L I P G D
TCTAAGGTGTACACATCGAATGTTCCGCTGTGTAGCATGCGAGCAGCTCATCCCGGAGAG 462
E F A L R E D G L F C R A D H D V V E R A S
GAATTGCGCTGCGGGAGGATGGGCTTTTCCGCGCGGACACCATGATGTAGTGAGAGGGGACG 528
L G A G D P L S P L H P A R P L Q M A A E P
CTAGGAGCTGAGACCTCTCAGTCCCTTGATCCAGCGCGGCTCTGCAATGGCAGCGGAGGCC 594
I S A R Q P A L R P H V H K Q P E K T T R V
ATCTCCGCTAGGACGCGCTCTGCGCGCGCAGCTCCCAACAGCGCGAGAGACCCAGGATG 660
R T V L N E K Q L H T L R T C Y A A N P R P
CGGATGTGCTCAACGAAGAGCTGCACCTTGGGAGCTGCTACGAGCCACCTCGGACCA 726
D A L M K E Q L V E M T G L S P R V I R V W
GATGCGCTCATGAAGAGCAATAGTGGAGATGACCGGCTCAGTCCCGAGTCTCCGGTCTGG 792
F Q N K R C K D K K R S I M M K Q L Q Q Q Q
TTTCAAAACAGAGGTGCAAGGACGAAGAACGACGATCATGTGAAGCAGCTCCAGCAGCAGCA 858
P N D K T N I Q G M T G T P M V A A S P E R
CCCAACGACAAAATAATATCCAGGGGATGACAGGAATCCCATGTTGGCTGTAGTCCGGAGAGA 924
H D G G L Q A N P V E V S Y Q P P W K V L
CATGATGTGTTTACAGGTACCCAGTTCAGGTGCAAGTACAGCCGCGCTCGGAAAGTACTG 990
S D F A L Q S D I D Q P A F Q Q L V N F S E
AGTGACTTCGCTTGAAGGTGACATAGATCAGCGCTGCTTTTCAGCACTGGTCAATTTTTCAGAA 1056

G G P G S N S T G S E V A S M S S * Q L P D T
GGAGGACGAGGCTTAATTCACCTGGCAGTGAAGTAGCATCGATGCTCTCAGCTCCGAGATACA 1122
P N S M V A S P I E A *
CCCAACGACATGGTACGAGTCTATAGAGGATGAGGAACATTCATCAGATGTTTGTGTTTGT 1188
TTGTTTTCCTCTGTTGGAGAACGTTGGAAATGACGTTGAACCTCCGAATAAAGTATTATACGAC 1254
CCAGTCAATGAAATCAATCAAGAACGATGCTCCATGAAATGCACGAGTCTGTTTAAATGAC 1320
AAGGTGATATGGTACCAACTGTGAAGACATCATGGGATTTTACTAGAAATTAACCAACACAAA 1386
ACCTTAAGCCCAACATATGCTATTCATGACCTTAGGAGGACTGAAAGAAAGGAAAGGAAAAA 1452
AGAGAGAGAGAGAGACGCTTTTAAACGTAAGAGATTATATTCAGGATCTCAAAACTGCGCA 1518
TTTTCATTTCACTGCACTATAGAGAAAGCAGAAATACAGAAATTCCTAGTCCATCTATTCTG 1584
AATGGTGTCTTTATATTTGGTCAATGCTTGCACAAACAGGAGCTCCAGCATAGAGCGGAAGAA 1650
CCAGCTCAGTGGCTGAAGAGTCTTTCAGGAAGCGGGGCTGCATGCTGCTGCTGTTTTCGCA 1716
TGTTTGTAGTACCTTGACAGAGGGTTAAGTGAATTTCTGGGTCTCTTAGCATGCGCTGTACCT 1782
GGTTCTCTTTTACCTTTGCTCTCTCCCATTAATTTCTTCTTTTCTTTTCTTTTCTTTTACAT 1848
TTTCTTGTAGATCATCTCTATCAAGAGCTGGAAGGACTTTAAAGTTTGTGAATTTATATTTA 1914
AAACCAACTTATAAGCATTGCAACAGGTTACCTCTATTTTGGCACAAGGCTCTCGGATTTGT 1980
TTTGACTGCTGCTGCTGCAAGAACTTTCCGCCAAGAGTGTATAGTATTGTTTAAATGACTG 2046
TTTTCGCTCTTTTCGAAATAAAGAGGAAAGAACTTTTGTGTTTGTCTTTCATTCGCA 2112
AAATTAATAAGTAAATTTATTTATTTATTTAGTCAGGAGACTTGCCACTTTTCATGTCATTG 2178
CTTTTGTGTTGCTGAAGTAAAAAGAAAGAGTTGTACCGTGGCTTTTGAATTAATATGTCTAA 2244
TTTATGTGTTTGTCTTTTCTTTTAAATATTTATGTGAATCAAGCGCATATGTGAATTAAT 2310
ATCTTCAGGACTATTTCACATAAAGCGTTTGGCATAGATAATTAATAAAAAAAGAAAAA 2376
AAAAAAGAAAAA 2389

FIG. 2 Nucleotide and deduced amino-acid sequence (single-letter code) of rat IsI-1 cDNA. The DNA sequence shown was derived from the combination of several different cDNAs, and the numbering of the nucleotides is shown to the right. The deduced amino-acid sequence starts with a putative initiation methionine. The homeodomain is indicated by a solid line, and the amino acids defining the LIM motifs⁵ are boxed. Glutamine residues in the C-terminal part of IsI-1 are indicated by asterisks.

METHODS. The RIN library was rescreened using the isolated 960-bp cDNA as a probe, and several recombinants isolated. Several genomic clones were also isolated and subfragments from the IsI-1 gene were used as probes to isolate cDNAs carrying copies of the 5' parts of the IsI-1 messenger RNA. The different cDNAs were combined to produce the sequence shown. The DNA sequence of the isolated cDNAs were determined on both strands using the dideoxy chain termination method⁴.

motif is mutated. There was no binding to a probe containing point mutations at positions -212, -217 and -219, respectively; each of the mutations is located in one of the three alternative TAAT motifs. There was specific binding to the corresponding sequence of the rat insulin II gene (Fig. 1b), which contains only one of the TAAT motifs. The wild-type rat insulin I and

II gene sequences competed for binding to the rat I sequence, whereas the S20 mutant did not (Fig. 1c). Taken together, these results support the involvement of the TAAT motif in binding.

We assembled a 2,400 base pair (bp) cDNA using overlapping cDNA clones, which contained a putative initiator methionine codon followed by an open reading frame encoding a poly-

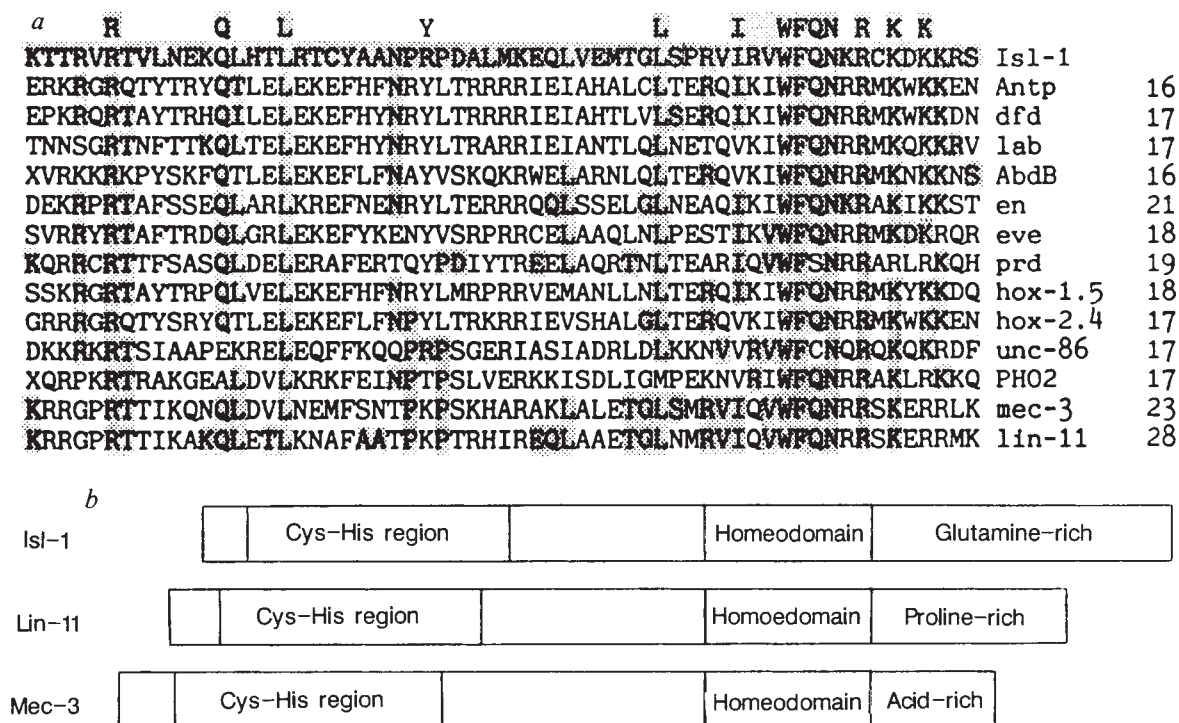


FIG. 3 **a**, Comparison of the Isl-1 homeodomain with representative members of the classes of homeodomain-containing proteins defined in ref. 12, and with the homeodomain of the *lin-11* gene product⁵. (Single-letter amino-acid code.) Amino acids identical to those of the Isl-1 homeodomain are shaded; numbers represent the number of identities found between the Isl-1 homeodomain and the other homeodomains. The 13 most-conserved homeodomain

amino acids are also indicated (top). **b**, Schematic representation of the overall structures of Isl-1, Mec-3 and Lin-11. The relative positions of the homeodomains and the Cys-His domains in Isl-1, Mec-3 and Lin-11 are shown. Glutamine-rich, proline-rich and acidic amino acid-rich domains reminiscent of known transactivation domains are also indicated.

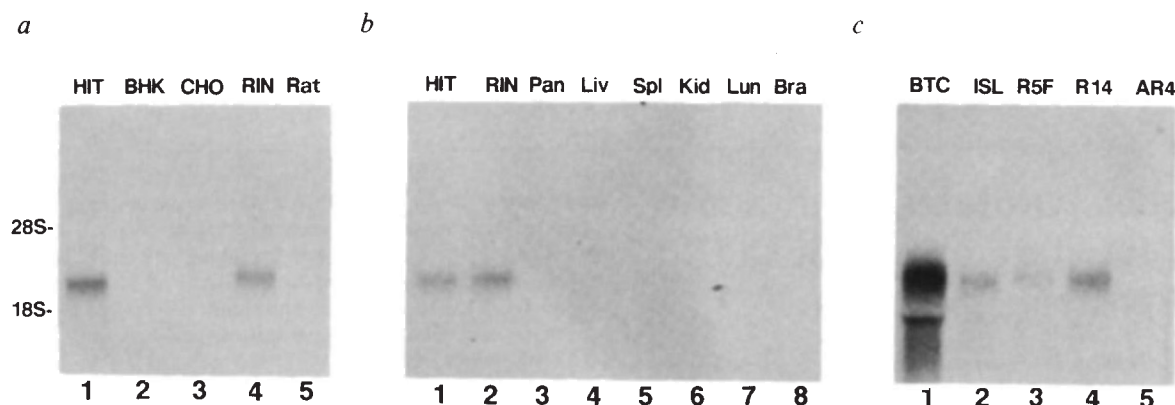


FIG. 4 Northern-blot analysis of Isl-1 RNA. **a**, Poly(A)⁺ RNA (7 µg per lane) from insulin producing HIT¹ and RIN 5F⁶ cells (lanes 1 and 4) and from noninsulin-producing BHK, CHO and Rat-1 cells (lanes 2, 3 and 5) was hybridized to an Isl-1 cDNA probe. The positions of 28S and 18S ribosomal RNAs are indicated. **b**, Analysis of poly(A)⁺ RNA (3 µg per lane) from various rat tissues. HIT and RIN poly(A)⁺ RNA were used as positive controls (lane 1 and 2). Lane 3, pancreas; lane 4, liver; lane 5, spleen; lane 6, kidney; lane 7, lung; lane 8, brain. **c**, Analysis of total RNA from purified rat islet cells and pancreatic cell lines. Lane 1, total RNA (30 µg) from β-TC cells¹⁵; lane 2, 15 µg total RNA from islet cells; lane 3, 30 µg RIN 5F⁶ RNA; lane 4, 30 µg RIN 14B⁶ RNA; lane 5, 30 µg RNA from the pancreatic exocrine cell line AR4-2J¹⁴.

METHODS. RNA from HIT, BHK, CHO and Rat-1 cells was prepared by the LiCl method¹⁶ and subjected to oligo(dT) affinity chromatography. Adult

normal rat islet cells were prepared as previously described¹⁷. Total-RNA was prepared using the guanidinium thiocyanate method¹⁸. Poly(A)⁺ RNA from rat organs was from Clontec laboratories Inc. RNA samples were electrophoresed through formaldehyde-agarose gels, transferred to Hybond-N (Amersham) and then crosslinked according to the manufacturer. The blot was probed with 3×10^6 c.p.m. ml⁻¹ ³²P-random-primed cDNA probe in 50% formamide, 5 × SSC, 5 × Denhardt's solution, 0.1% SDS, 10 mM EDTA, 10 mM Na₂HPO₄, pH 7.0, and denatured calf thymus DNA (100 µg ml⁻¹) at 42 °C. The blot was washed twice for 15 min in 2 × SSC, 0.1% SDS, and twice for 15 min in 0.2 × SSC, 0.1% SDS at 60 °C and exposed to Kodak XAR film for 24–48 h. The quality of all these RNA preparations was found to be comparable using actin cDNA probes, and for total pancreas and AR4-2J RNA, actin and trypsin cDNA probes.

peptide of relative molecular mass 38,000 (M_r 38K). The deduced amino-acid sequence of this polypeptide, denoted Isl-1, is shown in Fig. 2. Isl-1 contains two distinct domains; a homeodomain located between residues 180 and 240, and two copies of a finger-like Cys-His motif located in the N-terminal part of the protein. The homeodomain of Isl-1 has 12 out of the 13 most-conserved homeodomain amino acids but is only distantly related (25–30% identity) to all known *Drosophila* and mammalian homeodomains¹² (Fig. 3a). Instead, this homeodomain is most similar to the homeodomains of the *C. elegans mec-3* and *lin-11* gene products^{4,5}.

It is interesting that the homologies between these three proteins extend to include the Cys-His motifs, denoted the LIM motifs⁵; for this part, Isl-1 is even more related to Lin-11 than Mec-3 is to Lin-11. All three genes carry two tandem copies of the LIM motifs⁵, and, as outlined in Fig. 3b, the overall structures of these three proteins are very similar and together define a new class of homeodomain-containing proteins. By analogy to zinc-finger structures which are present in a variety of regulatory proteins (reviewed in ref. 13), the LIM motifs could coordinate metal ions stabilizing a specific structure of this domain, and, like other finger-like structures, act as a DNA-binding domain. It is also possible that the LIM and homeodomains of these proteins form a bipartite structure in which both domains contribute to the interaction with a single site. Alternatively, the LIM domain could interact with other proteins as a transcriptional activation domain, or interact with a protein(s) that regulates the activities or the subcellular localization of such proteins, or mediate dimerization. The availability of a cloned cDNA for Isl-1 should permit a detailed investigation of the function(s) of this domain.

We analysed the tissue distribution of Isl-1 RNA by northern-blot analysis. An Isl-1-specific transcript of about 3 kb was detected in poly(A)⁺ RNA from insulin-producing cell lines but not from heterologous cell lines or from a variety of adult rat organs including pancreas (Fig. 4, a and b). The lack of a detectable transcript in poly(A)⁺ RNA from total pancreas indicates an islet cell-specific expression of Isl-1, because islet cells only constitute about 1–2% of the total cell mass in pancreas. We purified normal adult rat islet cells and detected the Isl-1 transcript in 15 µg total RNA from the islet cells and in total RNA from a variety of cell lines of pancreatic endocrine origin (Fig. 4c). The islet-cell specificity of Isl-1 is also supported by the lack of the Isl-1 transcript in the pancreatic exocrine cell line AR4-2J (ref. 14; Fig. 4c).

Because Isl-1 is expressed in normal adult islet cells, it is likely that Isl-1 has a function in terminally differentiated islet cells. The expression of Isl-1 correlated with the expression of the insulin gene, with the possible exception of Isl-1 expression in a variant of the RIN cells (RIN 14B)⁶, which produce some insulin but predominantly somatostatin. But islet hormones are coexpressed during the normal embryonic development of endocrine pancreas¹⁹, and RIN cells coexpress most, if not all, islet hormones, and change their pattern of expression. It is therefore thought that RIN cells are islet stem cells or precursor cells²⁰. Therefore, the presence of Isl-1 RNA in these presumed islet stem cells indicates that Isl-1 is also expressed during development of islet cells. The presence of a homeodomain in Isl-1 also indicates a possible function for Isl-1 in the development of islet cells, particularly if the structural homologies to the proteins encoded by *mec-3* and *lin-11* also reflect functional similarities. The functions of *mec-3* and *lin-11* have been genetically defined: *mec-3* is needed for the specification of the touch neuron phenotype, and *lin-11* controls cell identities in some vulval precursor cells^{5,21}. By analogy, Isl-1 could be involved either in the specification of the β -cell phenotype or in the specification of the pancreatic endocrine cell lineage. □

Received 27 December 1989; accepted 9 February 1990.

- Karlsson, O. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 6453–6457 (1982).
- Karlsson, O., Walker, M. D., Rutter, W. J. & Edlund, T. *Molec. cell. Biol.* **9**, 823–827 (1988).
- Ohlsson, H., Karlsson, O. & Edlund, T. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4228–4231 (1988).
- Way, J. C. & Chalfie, M. *Cell* **54**, 5–16 (1988).
- Freyd, G., Kim, S. K. & Horvitz, H. R. *Nature* **344**, 876–879 (1990).
- Bhathena, S. J. *et al. Diabetes* **31**, 521–531 (1982).
- Singh, H., LeBowitz, J. H., Baldwin, A. S. Jr & Sharp, P. A. *Cell* **52**, 415–429 (1988).
- Klemm, H.-H. & Sippel, A. E. *EMBO J.* **6**, 2719–2725 (1987).
- Angel, P. *et al. Nature* **332**, 166–171 (1988).
- Odenwald, W. F. *et al. Genes Dev.* **3**, 158–172 (1989).
- Mihara, H. & Kaiser, E. T. *Science* **242**, 925–927 (1988).
- Scott, P. M., Tamkun, J. W. & Hartzell, G. W., III. *BBA Rev. Cancer* **989**, 25–48 (1989).
- Evans, R. M. *Science* **240**, 889–895 (1988).
- Stratowa, C. & Rutter, W. J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4292–4296 (1986).
- Efrat, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 9037–9041 (1988).
- Auffrey, C. & Rougeon, F. *Eur. J. Biochem.* **107**, 303–314 (1980).
- Lernmark, Å., Nathans, A. & Steiner, D. F. *J. Cell Biol.* **71**, 606–623 (1976).
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **5294**–5299 (1979).
- Alpert, S., Hanahan, D. & Teitelman, G. *Cell* **53**, 295–308 (1988).
- Madsen, O. D. *et al. J. Cell Biol.* **103**, 2025–2034 (1986).
- Ferguson, E. L., Sternberg, P. W. & Horvitz, H. R. *Nature* **326**, 260–267 (1987).

ACKNOWLEDGEMENTS. We thank Drs C. Betzholtz and M. Welsh for the gifts of the RIN λ gt 11 cDNA library and for some poly(A) and total RNA preparations from cell lines and adult islet cells, respectively, and Drs R. Horvitz and M. Chalfie for unpublished results. This work was supported by the Swedish National Science Research Council, the Swedish Medical Research Council and the Swedish National Board for Technical Development.

Heat-shock proteins DnaK and GroEL facilitate export of LacZ hybrid proteins in *E. coli*

Gregory J. Phillips & Thomas J. Silhavy

Department of Biology, Princeton University, Princeton, New Jersey 08544, USA

THE use of *lacZ* gene fusions, producing a hybrid protein containing an amino terminus specified by a target gene fused to the functional carboxy terminus of β -galactosidase, has facilitated the study of protein targeting in various organisms. One of the best characterized fusions in *Escherichia coli* is Φ (lamB–LacZ)42-1(Hyb), which produces a hybrid protein with the signal sequence and 181 N-terminal amino acids of the exported protein LamB, attached to LacZ¹. In common with other LacZ hybrids (reviewed in ref. 2), the LamB–LacZ(42-1) protein is poorly exported from *E. coli*, conferring a Lac⁺ phenotype. β -Galactosidase activity decreases markedly when cells producing the LamB–LacZ protein

are grown at 42 °C or when a heat-shock response is induced at lower temperatures by overproducing heat-shock factor RpoH³, indicating the LacZ hybrids are being efficiently targeted to the cell envelope. We now report that the heat-shock proteins DnaK and GroEL can, in sufficient amounts, decrease β -galactosidase activity and facilitate the export of lacZ-hybrid proteins.

High-level synthesis of the LamB–LacZ(42-1) protein (brought about by maltose induction) causes a lethal 'jamming' of the protein export machinery; a phenotype termed overproduction lethality². The LamB–LacZ(330) hybrid, however, contains only a small portion of lamB, is not exportable and does not confer overproduction lethality⁴; heat-shock induction has no effect on β -galactosidase activity of this hybrid.

In order to determine which proteins are responsible for decreasing β -galactosidase activity in the LamB–LacZ(42-1) fusion strains, the heat-shock proteins GroEL and DnaK were overproduced from plasmids, as described in Table 1. β -Galactosidase activity was decreased in cells overproducing either GroEL (lines 1, 2, 4, 5) or DnaK (lines 11–14) and, again, no effect on β -galactosidase activity was seen with cells containing the non-exportable hybrid protein LamB–LacZ(330) (lines 7–10, 15–18). SDS-PAGE⁵ showed that overproduction of the heat-

TABLE 1 β -Galactosidase activity of LamB-LacZ fusion strains

Strain	LamB-LacZ fusion	Plasmid	Maltose induction	β -Galactosidase activity (units/ A_{600})
1. GP3060	42-1	pBR322 (control)	-	174
2. GP3064	42-1	pOF39 (<i>groELS</i> ⁺)	-	39
3. GP3065	42-1	pOF13 (<i>groEL</i> ⁺)	-	52
4. GP3060	42-1	pBR322	+	846
5. GP3064	42-1	pOF39	+	352
6. GP3065	42-1	pOF13	+	314
7. GP3062	330	pBR322	-	292
8. GP3067	330	pOF39	-	297
9. GP3062	330	pBR322	+	1,335
10. GP3067	330	pOF39	+	1,350
11. GP3069	42-1	pGP120 (<i>dnaK</i> ⁻)	-	165
12. GP3070	42-1	pCG217 (<i>dnaK</i> ⁺)	-	0
13. GP3069	42-1	pGP120	+	824
14. GP3070	42-1	pCG217	+	22
15. GP3072	330	pGP120	-	294
16. GP3073	330	pCG217	-	279
17. GP3072	330	pGP120	+	1,371
18. GP3073	330	pCG217	+	1,284

Cells were grown in M63 medium¹⁹ supplemented with 0.05% thiamine, 0.4% glycerol, and 0.5% L broth²⁰ and appropriate antibiotic (ampicillin, 50 μ g ml⁻¹ for *groELS* and *groEL* plasmids and chloramphenicol, 10 μ g ml⁻¹ for *dnaK* plasmids). β -Galactosidase assays¹⁹ were performed on exponentially growing cultures at 30 °C. Enzyme activities are the average of at least two experiments. All strains are derivatives of MC4100²⁰ and contain either Φ (*lamB-lacZ*)42-1(Hyb) (signal sequence plus 181 amino acids LamB)⁴, or Φ (*lamB-lacZ*)330(Hyb) (signal sequence plus 20 amino acids of LamB)⁴. Maltose induction was performed in the presence of 0.4% maltose (1 h) before assay. Plasmids have been described previously: pOF39, pOF13 (ref. 21), pCG217 (ref. 22). pGP120 is a derivative of pCG217 constructed for this study by introducing a 432 basepair deletion between two *EcoRV* restriction sites within the coding region for *dnaK*¹⁶.

shock proteins did not significantly affect either the synthesis or stability of the hybrid proteins [LamB-LacZ(42-1) or LamB-LacZ(330)] (data not shown), further suggesting that changes in activity were the result of more efficient envelope targeting.

Direct evidence to support our hypothesis that the heat-shock proteins DnaK and GroELS improve export of the LamB-LacZ(42-1) hybrid protein was provided by observing increased efficiency of signal sequence cleavage. Processing is a reliable indicator that the amino-terminal portion of an exported protein has traversed the inner membrane. Figure 1 shows that following 1.0 h of maltose induction, only those cells overproducing GroELS or DnaK show significant signal sequence processing. To ensure that the results with DnaK depend upon the plasmid containing an intact *dnaK* gene, a large deletion was introduced into the *dnaK* open reading frame. As shown in Table 1 (lines 11-14) and Fig. 1, a functional DnaK protein is necessary to decrease the β -galactosidase activity and promote processing of the LamB-LacZ hybrid protein. GroELS is known to form a decatetramer complex consisting of two proteins⁶, GroEL⁷ and GroES⁸, of relative molecular mass 65,000 and 15,000 (*M*, 65K and 15K) respectively. Both proteins are encoded by the *groELS* operon⁸. To determine if overproduction of both components of GroELS is required for the observed effects on hybrid protein translocation, a plasmid which contains only an intact *groEL* gene was transformed into the *lamB-lacZ* fusion strain. Table 1 (lines 1-6) and Fig. 1 both show that overproducing GroEL alone was sufficient to promote LacZ hybrid protein targeting, albeit to a lesser degree than overproduction of DnaK.

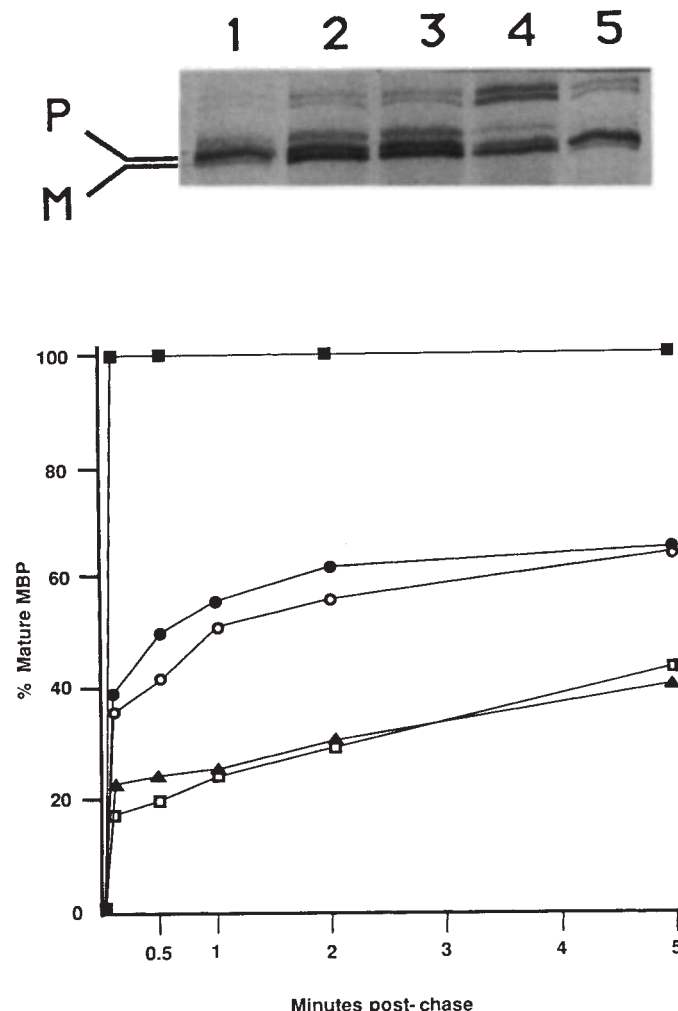


FIG. 1 Signal sequence processing of LamB-LacZ (Hyb42-1) protein. Lane 1, GP3060 (pBR322, control); lane 2, GP3064 (pOF39, *groELS*⁺); lane 3, GP3065 (pOF13, *groEL*⁺); lane 4, GP3070 (pCG217, *dnaK*⁺); lane 5, GP3069 (pGP120, *dnaK*⁻). P, precursor (unprocessed) form; M, mature (processed) form.

METHODS. Cultures were grown as described in the legend to Table 1. At an absorbance reading at 600 nm of 0.200, maltose was added to the cultures to a final concentration of 0.4%. After continued incubation (1 h) the cells were pulse-labelled with 50 μ Ci ml⁻¹ of Tran³⁵S-Label (ICN), for 30 s, and chased with an equal volume of prewarmed, supplemented M63 medium containing 0.8% unlabelled methionine. After a 5-min chase, a 1-ml aliquot of cells was removed, precipitated and washed as described²³. Samples were dissolved in 50 μ l protein sample buffer (0.5 M Tris-HCl, pH 6.8, 10% glycerol, 2% β -mercaptoethanol, 2% SDS, 1% Triton X-100, 0.0025% bromophenol blue) and boiled before electrophoresis in a 6% polyacrylamide gel⁵ using a Mini-protein II dual slab cell (Bio-Rad). Only the relevant portion of the gel is shown. The top two bands represent the β - and β' -subunits of *E. coli* RNA polymerase.

FIG. 2 Kinetics of signal sequence processing of maltose-binding protein following maltose induction. GP3060, (filled triangles); GP3064, (open circles); GP3065, (filled circles); GP3070, (filled squares); GP3069, (open squares).

METHODS. Cells were grown, maltose-induced and pulse-labelled (see Fig. 1 legend). Following pulse-labelling, 0.5 ml samples were removed at 10 s, 0.5, 1, 2, and 5 min after addition of chase mixture. MBP was immunoprecipitated using rabbit antiserum²³ before electrophoresis on an 8% polyacrylamide gel⁵. Autoradiographs were analysed on a Hoeffer GS300 transmittance/reflectance scanning densitometer. Peaks were integrated using the Hoeffer Scientific GS350 program.

The question as to the specific location of the hybrid protein within the cell envelope is under investigation, but it appears to be membrane-associated.

If heat-shock proteins increase the export efficiency of the LacZ hybrid protein, we predicted decreased jamming of the protein export machinery caused by high-level synthesis (maltose induction) of LamB-LacZ(42-1). To test this prediction we followed the kinetics of signal sequence cleavage of maltose-binding protein (MBP) after maltose induction. As shown in Fig. 2, processing of the MBP signal sequence in control cells is almost completely blocked after 1.0 h of maltose induction. Yet in cells overproducing functional DnaK, MBP processing remains extremely efficient. Overproducing GroEL was likewise sufficient to maintain export of MBP, although not as efficiently as DnaK overproduction.

GroEL is a member of the heat-shock protein hsp70 family of molecular 'chaperones'⁹ and evidence has been presented to suggest that it may participate in bacterial protein export by maintaining polypeptides in a translocation competent conformation¹⁰⁻¹². Our results support and extend this suggestion. Hsp70 proteins appear to function in similar fashion during protein translocation in eukaryotic cells¹³⁻¹⁵. DnaK is a member of this protein family¹⁶ and we show that it can also function in bacterial protein export. Moreover, as GroEL and DnaK are overproduced at comparable levels with the plasmid constructs described, we conclude that DnaK is far more efficient in this regard than is GroEL.

Given our results, it seems likely that the problems associated with LacZ hybrid export are the result of protein folding¹⁷. DnaK and GroEL exert their effects by allowing the molecules to retain a conformation compatible with the protein export machinery. The discovery that the hsp60 and hsp70 proteins

function in this regard in *E. coli* provides a convenient way to study their mode of action¹⁸. For example, by scoring LacZ activity of gene fusions such as $\Phi(lamB-lacZ)42-1(Hyb)$, we can further analyse the effects on protein export and polypeptide folding by this important class of proteins. □

Received 19 January; accepted 7 March 1990.

1. Silhavy, T. J., Shuman, H. A., Beckwith, J. & Schwartz, M. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5411-5414 (1977).
2. Bleker, K. L., Phillips, G. J. & Silhavy, T. J. *J. Bioenergetics Biomembranes* (in the press).
3. Grossman, A. D., Straus, D. B., Walter, W. A. & Gross, C. A. *Genes Dev.* **1**, 179-184 (1987).
4. Benson, S. A., Bremer, E. & Silhavy, T. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3820-3834 (1984).
5. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
6. Chandrasekhar, G. N., Tilly, K., Woolford, C., Hendrix, R. & Georgopoulos, C. *J. biol. Chem.* **261**, 12414-12419 (1986).
7. Georgopoulos, C. P. & Hohn, B. *Proc. natn. Acad. Sci. U.S.A.* **75**, 131-135 (1978).
8. Tilly, K., Murialdo, H. & Georgopoulos, C. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1629-1633 (1981).
9. Hemmingsen, S. M. *et al. Nature* **333**, 330-334 (1988).
10. Bochkareva, E. S., Lissin, N. M. & Girshovich, A. S. *Nature* **336**, 254-257 (1988).
11. Lecker, S. *et al. EMBO J.* **8**, 2703-2709 (1989).
12. Kusakawa, N., Yura, T., Ueguchi, C., Akiyama, Y. & Ito, K. *EMBO J.* **8**, 3517-3521 (1989).
13. Deshaies, R. J., Koch, B. D., Werner-Washburne, M., Craig, E. A. & Schekman, R. *Nature* **332**, 800-805 (1988).
14. Chirico, W. J., Waters, M. G. & Blobel, G. *Nature* **332**, 805-810 (1988).
15. Murakami, H., Pain, D. & Blobel, G. *J. Cell Biol.* **107**, 2051-2057 (1988).
16. Bardwell, J. C. A. & Craig, E. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 848-852 (1984).
17. Lee, C., Li, P., Inouye, H., Brickman, E. R. & Beckwith, J. *J. Bact.* **171**, 4609-4616 (1989).
18. Rothman, J. E. *Cell* **59**, 591-601 (1989).
19. Miller, J. *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratories, Cold Spring Harbor, New York, 1984).
20. Silhavy, T. J., Berman, M. L. & Enquist, L. W. *Experiments with Gene Fusions* (Cold Spring Harbor Laboratories, Cold Spring Harbor, New York, 1984).
21. Fayet, O., Louarn, J.-M. & Georgopoulos, C. *Molec. gen. Genet.* **202**, 435-445 (1986).
22. Zyllics, M. & Georgopoulos, C. *J. biol. Chem.* **259**, 8820-8825 (1984).
23. Stader, J., Benson, S. A. & Silhavy, T. J. *J. biol. Chem.* **261**, 15075-15080 (1986).

ACKNOWLEDGEMENTS. We thank Costa Georgopoulos and Carol Gross for plasmids, and members of the Silhavy laboratory for helpful suggestions. G.J.P. is supported by a postdoctoral fellowship from the American Cancer Society, T.J.S. by a National Institutes of Health grant.

GUIDE TO AUTHORS

Please follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers for whom English is a second language and those who work in other fields. Please write clearly and simply, avoiding unnecessary technical terminology. *Nature's* staff will edit manuscripts to those ends if necessary. Contributors should check their proofs carefully.

Because of the competition for space, many of the papers submitted for publication cannot be accepted. For this reason, and because brevity is a great assistance to readers, papers should be as brief as is consistent with intelligibility. Please note that one printed page of *Nature*, without diagrams or other interruptions of the text, has fewer than 1,300 words.

CATEGORIES OF PAPER

Review Articles survey recent developments in a field. Most are commissioned, but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance but must not exceed six pages of *Nature*.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text or more than six display items (figures and tables) and should not occupy more than five pages of *Nature*.

Articles should be accompanied by a heading of 50-80 words written to advertise their contents in general terms, to which editors will pay particular attention. A heading (printed in italic type) is not an abstract and should not usually contain numbers or measurements. The study should be introduced in more detail in the first two or three paragraphs, which should also briefly summarize its results and implications.

Articles may contain a few subheadings of two or three words. The meaning of the text should not depend on the subheadings, whose function is to break up the text and to point to what follows. There should be fewer than 50 references.

Letters to Nature are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should not have more than 1,000 words of text or more than four display items and should not occupy more than two pages of *Nature*. The first paragraph should describe, in not more than 150 words, the origins and chief conclusions of the study. Letters should not have subheadings or more than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned, most are unsolicited. They are normally between one and four pages of *Nature* in length.

News and Views articles are intended to inform non-specialist readers about a recently published advance. Suggestions should be made to the News and Views Editor. Illustrations are welcome. Proposals for meeting reports should be agreed in advance.

Scientific Correspondence is for the discussion of scientific matters, including contributions published in *Nature*. Priority is given to contributions of less than 500 words and five references. Figures are welcome.

Submission. All manuscripts may be submitted to either London or Washington. They should not be addressed to editors by name. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax (VAT).

GENERAL

All contributions submitted for publication in *Nature* should conform with these rules.

Manuscripts should be typed, double-spaced, with a good-quality printer, on one side of the paper only. Four copies are required, each accompanied by lettered artwork. Four copies of half-tones should be included. Reference lists, figure legends and so on should be on separate sheets, all of which should be double-spaced and numbered. Copies of relevant manuscripts in press or submitted for publication elsewhere should be included, clearly marked as such. Revised and resubmitted manuscripts should also be clearly marked as such and labelled with their reference numbers.

Titles should say what the paper is about with the minimum of technical terminology and in fewer than 80 characters. Authors should avoid active verbs, numerical values, abbreviations and punctuation and should include one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name, figure number and, when known, manuscript reference number. Original artwork should be unlettered. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are permitted only if the parts are closely related, either experimentally or logically. Suggestions for cover illustrations, with captions, are welcome. Original artwork (and one copy of the manuscript) will be returned when a manuscript cannot be published.

Colour Artwork. A charge of £500 a page is made for 4-colour figures. Indentation to meet these costs will not prevent the publication of essential colour figures if the circumstances are explained.

Figure legends must not exceed 300 words and ideally should be much shorter, and should use telegraphic form wherever possible. The figure should be described first, then, briefly, the method. Reference to a method described elsewhere is preferable to a full description. Methods should not be described in detail in the text.

References should be numbered sequentially as they appear in the text, followed by those in tables and finally those in the figure legends. Reference numbers apply only to papers published or in the press, and a different number should be given to each paper cited. All other forms of reference (including unreferenced abstracts) should be included in the text as personal communication, manuscript in preparation or preprint (with number and institution where appropriate). Text should not be included in the reference list. References should be abbreviated according to the *World List of Scientific Periodicals* (fourth edition, Butterworth, London, 1963-65). The first and last page numbers should be cited. Reference to books should clearly indicate the publisher and the date and place of publication.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided as much as possible and, when used, defined. Editors will shorten words if necessary. In case of doubt, SI units should be used.

Footnotes should not be used except for changed addresses or to identify the corresponding author if different from the first-named.

Acknowledgements should be brief. Grant numbers and contribution numbers are not allowed.

Copies of articles from this publication are now available from the UMI Article Clearinghouse.

For more information about the Clearinghouse, please fill out and mail back the coupon below.

UMI Article Clearinghouse

Yes! I would like to know more about UMI Article Clearinghouse.

I am interested in electronic ordering through the following system(s):

- ☐ DIALOG/Dialorder ☐ ITT Dialcom
☐ OnType ☐ OCLC ILL Subsystem

☐ Other (please specify) _____

☐ I am interested in sending my order by mail.

☐ Please send me your current catalog and user instructions for the system(s) I checked above.

Name _____

Title _____

Institution/Company _____

Department _____

Address _____

City _____ State _____ Zip _____

Phone (_____) _____

Mail to: University Microfilms International
300 North Zeeb Road, Box 91 Ann Arbor, MI 48106

This ever-changing world

Michael Grubb

Environments at Risk: Case Histories of Impact Assessment. By Derek Ellis. *Springer-Verlag*: 1989. Pp. 329. Pbk DM68, £24.50, \$39.50.

Technology and Environment. Edited by Jesse H. Ausubel and Hedy E. Sladovich. *National Academy Press*: 1989. Pp. 221. \$35, £42.

Global Climate Change: Human and Natural Influences. Edited by S. Fred Singer. *Paragon House*: 1989. Pp. 424. \$34.95.

The Next One Hundred Years: Shaping the Fate of Our Living Earth. By Jonathan Weiner. *Bantam Books*: 1990. Pp. 312. \$19.95.

The End of Nature. By Bill McKibben. *Viking*: 1990. Pp. 212. £12.99.

THE greening of both politics and science has spawned many books. In ranging from outright scepticism to apocalyptic visions, this collection throws fascinating light on the role — and limits — of science in the environmental debate.

Ellis's *Environments at Risk* is a modestly conceived book focusing on localized environmental impacts, and is remarkably informative and objective. Careful accounts of past issues range from the centuries-old yet continuing story of monitoring and cleaning the River Thames, through some of the more recent legends of environmental history such as the Minamata tragedy and the *Amoco Cadiz*, up to Bhopal and Chernobyl. Many lesser-known cases are covered. The accounts are not florid tales of disaster, but studies of what happened and why, and the lessons which might be learned — including the role of media and public reaction. The studies are drawn as far as possible from primary sources — something which Ellis emphasizes as

critical in a field where derivative accounts often compound errors.

Accounts of disasters are always more compelling and easier to write than those of environmental successes, but Ellis is also at pains to cover some major industrial developments which avoided significant impact through careful planning. Industrial planners in particular may find these accounts the most informative. For Ellis is clearly of the view that industry can develop without severe environmental or social conflicts — but only if it proceeds with care and sensitivity.

Environments at Risk is not a fashionable book. The account of acid rain is among the weakest of the studies, and other hot topics of stratospheric and tropospheric ozone, waste and global warming are absent. The book is much too detailed and dry to make a popular read. But in addition to forming a valuable reference on many of the skeletons in the environmental cupboard, this volume should be a compulsory read for

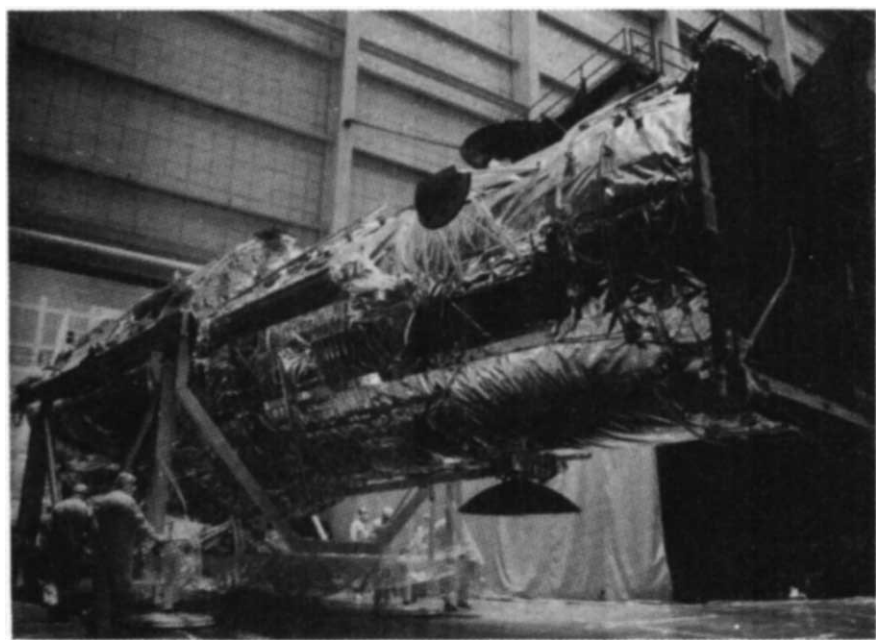
industrial environment managers.

Technology and Environment is a collection of essays which present a more academic and analytical assessment of broader environmental issues. The first three essays set the scene. "Industrial Metabolism" presents the broad concept of analysing industrial activity in terms of its overall materials and energy input, transformation and output — something which emphasizes the extraordinary wastefulness of industrial systems compared with any natural metabolism. As if unconsciously reinforcing this, "Dematerialism" casts doubt on the idea that advanced societies are tending to reduce materials consumption, pointing to various contrary trends that seem to offset specific technological improvements. "Regularities in Technical Development" cites evidence for patterns in technology deployment, especially in transport and energy sources. All three chapters are thought-provoking but ultimately frustrating, presenting many maybes but few hard conclusions beyond the obvious.

Part II, "The Promise of Technical Solutions", presents three competent but unremarkable essays, two on energy (primarily electricity) technologies, and one on chlorofluorocarbons. Far more interesting is part III, "Social and Institutional Aspects", which echoes the frustration of engineers: frustration at the constraints of the environment, the uncertainties of its demands, the ineptitude and inefficiency of environmental regulations, and the overall "paradox of technology development", in which engineers meet one set of demands only to be confronted with a dozen dilemmas from the side effects. In harmony with Ellis, all these essays point to the same conclusion: environmental confrontation patched up with end-of-pipe fixes cannot go on. If technology is to remain more cure than curse, industry needs to anticipate environmental issues and incorporate environmental objectives at the earliest stages of design and planning.

Technology and Environment is an attempt by engineers to develop some of the necessary understanding of broad technological themes. Unfortunately, the quality is patchy and the editing uneven — for example, the logistic curve of market penetration is used repeatedly, but a semi-rigorous description and explanation crops up only in one of the later chapters. The book contains nuggets but, given the depth of expertise included, it is disappointing. If it illustrates one thing, it is that assessing the social and technological side of the environmental equation is at least as difficult and uncertain as assessing the impacts.

But it is when environmental themes are taken to the global level that the going really gets messy. The collection of essays in *Global Climate Change* attempts to set



THE Hubble space telescope in May 1986 during its construction. From *The Space Telescope*, a study of NASA science, technology and politics, by Robert W. Smith and published by Cambridge University Press: 1990 (£40, \$39.50). At the time of going to press, the telescope is still on the launch pad. See the Commentary by Fred Hoyle on page 807 of this issue.

the biggest of them all in a sound scientific setting — and is hopelessly muddling. Refreshingly, the book does not equate climate change with atmospheric pollution, but runs the whole gamut: climatic history; volcanoes and meteorites; the hydrosphere issues of ocean, acid rain and desertification; and 'nuclear winter'. When there is controversy (which means almost everything) the lead paper is followed by a critique from another author. It should be a great book. Somehow, it isn't.

Part of the problem is structure and inadequate editing. Although the book starts with an informative overview essay on climatic history, modelling and various interpretations, it then dives straight into the greenhouse controversy. The "comment" by Ellsaesser is a full-blooded attack on both the institutional process of "consensus science" in the environmental area and on the specifics of greenhouse predictions, concluding that "climate models . . . generally overstate the climatic effect of doubled CO₂ by at least 2- to 3-fold". A reply to the comment seems to decimate Ellsaesser's arguments as dated misrepresentations of modern three-dimensional climate modelling, but ignores the rest of Ellsaesser's case. A short "where do we stand?" paper is then supposed to close the debate by reminding readers that it really is all rather complicated. Similar patterns emerge for ozone depletion, the nuclear winter (or summer?) debate, and volcano impacts.

This is a mixed volume. Some essays are attempts at balanced reviews; some are challenges; some are full-scale research papers. Most are discursive, a couple highly mathematical. Paradoxically, in a volume devoted to debate on physical climate issues, perhaps the best piece is the concluding stand-alone overview of Gaia.

The editor is clearly a universal sceptic, and gives the sceptics on each issue good space — and usually the last word. This gives a refreshing perspective and some challenging observations — but there are times when the scepticism seems to outweigh the science. A process which should lead to an appreciation of where the main uncertainties lie, and what are reasonable balances of probabilities, leads instead to an uneasy feeling that authors are vying for the reader's allegiance between orthodox and non-conformist parties — with the illogical implication that we might as well carry on changing the planet until we find out what we are doing.

Changing the planet is the more explicit theme of Weiner's *The Next 100 Years*. Rarely have I read a book with such a sharp contrast between the start and the core. After wading through half a dozen chapters of very basic science and at times laborious chat about various personal histories in climate observation, I was ready to give up. I'm glad I didn't. The science is

not first class — some questionable studies are uncritically reported and the discussion is not fully balanced. But the book is broad, and in this field at this time that is just as important. In attempting to make the science of climate change and impacts readable, Weiner succeeds in leaving a deep and powerful impression of the global environmental picture, and the many known and countless unknown ways in which humanity is interfering with it.

In "The Severn Spheres", for example, he describes some of the countless interactions and feedbacks between the different spheres of the globe. "Lovejoy's Islands" is a graphic account of how complex ecosystems can unwind when they are reduced below a certain size or overstressed, pointing out that at the same time as warming the climate, humanity is imposing all kinds of other stresses and chopping up the surface in ways which could make the migration of many species that follow the changing weather impossible. "The Oracle of Gaia" broadens the perspective further.

It seems that Weiner knows many of the key scientists well, and this is his homage to their work. He may have overdone the personalization and anecdotes, but the final product reflects the kind of broad intuition for the subject which can be gained only from such an association.

It is a pity that McKibben's book, *The End of Nature*, has attracted much more attention and will probably be much more famous. This is inevitable, not because McKibben is a doomsayer, but because of the book's title, prose and the fact that it takes one simple idea and elaborates upon it endlessly. The idea is not that we are about to destroy the world, but that humans are now affecting the biosphere on a global scale, and that therefore, "we have ended the thing that has, at least in modern times, defined nature for us — its separation from human society". McKibben is not saying that nature will end, but is defining it in such a way as to conclude that it has already ended.

McKibben makes a reasonable defence of the basic idea, and in the rest of the book he attempts to struggle with its implications. Much of the book is concerned with feelings, and we are treated to repeated, lengthy accounts of the forests where McKibben lives, with musings on the beauty of the world and the pity that in the author's sense it is no longer "natural". There is a half-hearted attempt to grapple with the philosophical and quasi-religious implications — after all, the idea does place humans in the position usually reserved for gods — but McKibben is no deep philosopher or theologian.

In considering the future, McKibben castigates "the defiant reflex" of those who think that humanity should in some way manage the world. In doing so he mixes up to an extraordinary degree the

terrifying simplicity of engineered solutions (such as putting dust in the stratosphere) with the efforts of ecologists seeking to minimize and ameliorate the impact of human activities on the biosphere to make reasonable trade-offs between human desires and impacts. Genetic engineering comes in along the way: "the idea that nature — that anything can be defined will soon be outdated. Because anything can be changed . . . 'Rabbit' will be a few lines of code. . ."

McKibben's alternative to the "artificial world" is the "humble idea" of "deep ecology", in which humans avoid all activities that could interfere in any way with the global system. Not merely caution, not dams in place of fossil fuel, but abstinence based on a universal non-human-centred philosophy. "Ecotage" — modern-day luddism — comes close to approval. McKibben notes that deep ecology "is an extreme solution", but "we live in extreme times". Nevertheless, even he shies away from a real examination of what this would mean or, much more important, how societies might get there. I can think of several paths but all of them seem markedly worse than facing up to a minimal degree of ecological management. By refusing to talk about trade-offs, McKibben is setting up a straw man against a dangerous intellectual vacuum.

It is clear from all these books that our understanding of almost everything to do with "the end of nature", to use McKibben's evocative phrase, is still desperately limited. Even the most basic questions remain unanswered, or even unasked. If not Gaia, in its broad sense of biological feedbacks stabilizing the Earth's system, then how has the present balance — so far from the physical equilibrium of a dead planet — been achieved and maintained? But if Gaia, then why the large swings of the ice ages in response to fairly minor orbital perturbations? How can we understand the reactions and limits of the global environmental system when we have no real understanding of how it works or evolved? And it is not only in the biosphere that our knowledge is so limited; as all the attempts to grapple with the societal causes and possible responses make clear, our understanding of the human side is just as poor. These books may not be great, but they are all significant contributions to the many-fold infant disciplines of global environmental sciences. □

Michael Grubb is at the Royal Institute of International Affairs, 10 St James's Square London SW1Y 4LE, UK.

■ In *The Greenhouse Effect: Negotiating Targets*, Michael Grubb examines the debate on international negotiations for limiting emissions of greenhouse gases and addresses the options and implications of failure to reach an international agreement. Available from the above address, price £10, \$20.

Balancing act

David Owen

The End of the American Century. By Steven Schlossstein. *Congdon and Weed*: 1989. Pp. 537. \$22.95. Distributed in the United Kingdom by Pandemic, £16.35.

THIS is a difficult book to read in as much as its style alternates between breathless journalism and detailed academic analysis. It starts with a chapter looking back from the year 2001 and asserts that Mikhail Gorbachev and his loyalists had been purged in 1992 and sent to Siberian camps, and that Japan had defeated the Soviet Union in a major military encounter involving the disputed Kurile Islands. The only justification for this chapter can be to jerk the reader out of any complacency about the future facing the United States. Yet its likely effect is to make the reader question the author's judgement.

The impression of superficiality is enhanced by the author's tendency to take a trip around the world with a mixture of pen portraits and quotes more appropriate to *Time* or *Newsweek*. The Office of Technology Assessment of the US Congress has undoubtedly published an important report called *International Competition and Services*. But it adds nothing to know that its project director is as comfortable "discussing disaggregated tax incentives as he is analysing trade and investment data" and is "tall and lean, with angular features and a thick, full beard," looking "very much like a youthful Abe Lincoln". Nor is it relevant that a young British foreign service officer in Singapore had clear blue eyes and a penetrating look that reminded him "very much of Robert Redford".

Yet for all the stylistic extravagancies, this book's 480 pages contain fascinating facts and figures comparing the economies of Asia with the United States which are not easily found in a single volume. Nor can the analysis of the lack of competitiveness of the United States be shrugged off as exaggeration.

Schlossstein's underlying theme is challenging but highly questionable. He argues that "stability and prosperity need more than just military might: they require an effectively functioning international financial system anchored by one dominant industrial nation whose currency is strong and commands respect". In as much as this was provided by Pax Britannica and the pound sterling and Pax Americana and the dollar during the past 150 years, Schlossstein feels the need to raise the spectre of Pax Nipponica and the yen. He writes that "Pax Americana can be revived, revitalised and restored. It must be. Because the alternatives,

whether an unpredictable dominance and control by Japan under Pax Nipponica or chaos and uncertainty resulting from a world with no dominant single power at all, are both less tolerable."

Pax Americana, Pax Nipponica, or with no Pax, anarchy. How simple the author makes it all sound. Yet what superficial nonsense it is. First, to view the past 150 years as a period of stability and prosperity is stretching one's sense of historical truth. Two world wars and innumerable conflicts can hardly be depicted as a stable



A world run by money?

period in world history. As for prosperity, it is too early to make an overall judgement. The pattern of prosperity has been very uneven. The depression in the 1930s cast a long shadow over the industrialized Western democracies and the sum total of human happiness in big areas like China, the Soviet Union and in the Indian sub-continent may well have been no greater in the past 150 years. We may have sown the seed of the destruction of our global environment; only time will tell us the answer. The author might be right that "anarchy may rule" with no Pax. But a coherent argument can be mounted for a multipolar world being more peaceful and prosperous.

Certainly, a multipolar world is a more attractive objective for most of the world's citizens than reinvigorating Pax Americana. Not many want to recycle Vietnam and all that came to represent. The Soviet Union, having given up its empire from economic weakness and now facing its own dismemberment, is no longer a superpower. Without such an overt Communist threat it is very unlikely that even an economically rejuvenated United States of America will want to continue to project its military strength as a superpower. The United States can and should improve its competitiveness and reverse some of its

declining industrial sectors, but it will do this for its own internal reasons.

The strength of this book is in Schlossstein's description of what is happening to the other economies in South-East Asia in South and North Korea, Taiwan, Hong Kong, Singapore and the Philippines. In effect these countries are likely to be a considerable check on Japan. If the Chinese coastal strategy for economic development works, that too will counter Japanese dominance. It is quite likely therefore that Japan's position will mirror that of a united Germany in Europe, both countries constrained and influenced by their immediate neighbours. Of course in Europe this process is institutionalized in a sophisticated way within the European Community, itself likely to enlarge from 12 to 16 countries with the intergration of East Germany and the addition of Austria, Hungary, Czechoslovakia and Poland. In Asia there is only as yet ASEAN which, though in its infancy, has the potential to develop and to establish even closer links with Japan.

The dominance of a single currency is neither desirable nor likely. The deutschmark, the yen and the dollar will probably stay as the basic basket of currencies. For monetary dominance is not something national governments any longer aspire to because it makes them too vulnerable. A single currency for Europe, the ecu, is likely to emerge and although the yen may push the dollar aside in Asia it is not in the Tokyo financial market's interest to destroy Wall Street. Spreading the risk and the speculation in currencies is part of a global market.

It will be clear from my review that readers tempted by publishers' hype to believe that they are reading a work comparable to Paul Kennedy's *Rise and Fall* will be disappointed. Nor has this book got the bite or tautness of Jean-Jacques Servan-Schreiber's *The American Challenge*. It is nevertheless worthwhile dipping into on a long intercontinental air-flight or on vacation when tired of novels.

For a more compelling prescription of what the United States should do now to overcome its gigantic problems, its decaying infrastructure, inadequate public education, insufficient public revenue and excessive federal deficits, one could do no better than read Felix Rohatyn's article on *perestroika* for the United States in *The New York Review of Books* issue of 12 April, "Becoming what they think we are". The US problem is that the Democratic party in opposition is still treating active government and tax increases as taboo subjects. Until a real debate with President Bush's administration is provoked on these subjects, lessons about competitiveness from the Pacific rim will continue to fall on stony ground. □

David Owen MP is at the House of Commons, London SW1A 0AA, UK.

Soft money

Daniel S. Greenberg

A Fragile Power: Scientists and the State. By Chandra Mukerji. Princeton University Press: 1989. Pp. 253. \$24.95.

IT IS late to discover that the cohabitation between science and the US government is loveless and rooted in mutual exploitation. Or that obtaining grants "is hardly guaranteed in the system, and being able to support oneself as a researcher in the United States is not a simple job". Chandra Mukerji, professor of sociology and communication at the University of California, San Diego, has examined one sector of the US scientific enterprise, oceanography, and is dismayed to find a marriage of convenience with the federal government. Although her researches were focused on oceanographers, she claims her findings extend to other sciences, a tenuous proposition given the Navy's consuming interest in the ocean sciences. In any case, the result is what she describes as "a book that many scientists may find offensive because it is designed to attack precisely those areas of science and science policy that polite people have learned to ignore".

The book's weighty bibliography indicates that these areas are far from ignored. Moreover, it is doubtful that many scientists, polite or not, will find this volume offensive, let alone factually eye-opening. More likely, many will find it naive, but possibly also provocative and tantalizing for its fresh, though unfortunately poorly developed, assertions about the nature of relations between US researchers and their government patrons.

Central to Mukerji's thesis is a commonplace underpinning of US research economics, 'soft-money' grants, the scientific version of junk bonds — rewarding but risky. Little explored by other scholars of science policy, soft money is viewed by the author as a cunning mechanism for political management of science. Why do federal research agencies deal in soft money, which is of relatively short duration and uncertain renewability? Because "Soft money scientists constitute a reserve labor force in the sense that they are supported by government and industries so their honed skills will be available when they are needed (by, for instance, the military in case of war, by industry in case there are major changes in the direction of the economy, or by the medical community if there is an outbreak of some new and threatening illness)".

What does government want of these scientists? Based on interviews with oceanography and their federal managers and policy makers, Mukerji states that

"government interest in research was not well explained by government need for the *information* [italics in original] generated by scientific projects. I was led to the conclusion that when government agencies dispense money to scientists for research, they do it not so much because the state is interested in maximizing its store of scientific information relevant to policy issues, but more because the government has interests in maintaining a *labor force* of skilled scientists available for consultation on policy issues".

In reality, however, relatively few scientists are summoned to consult on policy issues. And of those called, many are financed firmly by industry or academe, or are tenured civil servants, in positions mercifully remote from the perils of soft money. Mukerji does not mention the burdens of the alternative to soft money, the eternal block grant, which arrives even after scientific senescence has set in.

Mukerji contends that "like the unemployed on welfare, scientists on research grants are kept off the streets and in good health because of the interests and investments by elites". They thus comprise a "reserve labor force" for meeting national needs, she says. But she rests this iconoclastic interpretation on the dubious asser-

tion that scientists, at least the outstanding ones, command welfare because "Almost any [other] government would welcome and support them". Now, the mandarins of science in the United States have used many rousing ploys to keep the money flowing — cold war, national prestige, economic competitiveness, cultural value — but flight abroad is not seriously in evidence among them.

A Fragile Power is strongest in its analyses of work at sea, particularly the operations of the deep-sea research submersible *Alvin*. Through interviews and narrative, Mukerji outlines the interplay of military and scientific considerations in deep-sea research, as well as the different interests and rivalries of the scientific disciplines involved. Her analyses of taped conversations from the *Alvin*'s studies of hydrothermal vents provide insights into the processes of scientific observation. And drawing on her studies of scientists at work, she correctly observes, in an unusually succinct passage, that "the government profits from the relative autonomy of scientists because scientists' belief in their own detachment helps give the voice of science the legitimating power that makes science a resource for the state".

But all too often, Mukerji fills space with trite statements, such as "Many research problems that are funded by government organizations are selected for their salience to clear practical problems", and "The government obviously does not want to fund work that is not good". Correct in both cases, and fortunately so.

Some passages are mysteriously dense: "For many younger and journeyman researchers, the time to advise the state is imminent. They do not feel they can preserve the purity of their research careers by seeking exclusively NSF funds. They look to the Department of Defense and the Department of Energy (among others) to make their research lives either possible or easier. They gain what autonomy they can within applied projects (if they care) and use what techniques they can either to keep publishing (and improving their cultural capital as scientists) or to normalize their advisory work (making themselves contented civil servants with jobs they enjoy doing)."

Mukerji states in her preface, "this is not the book I started out to write", explaining that she was diverted on discovering "how much of the life of the laboratory was devoted to generating or justifying research funds". Indeed it is so devoted, and the subject merits continuing examination. Regrettably, this offering will neither inform the cognoscenti of science nor illuminate its practices for outsiders who wish to understand. □

Daniel S. Greenberg is editor and publisher of *Science & Government Report*, 3736 Kanawha St NW, Washington, DC 20015, USA.

Spring paperbacks

■ A collection of paperbacks that would be the envy of anyone's holiday suitcase is issued this spring by Penguin. Published today is *What Mad Pursuit*, Francis Crick's autobiography, reviewed in *Nature* (336, 268; 1988) by John Cairns, price £4.99. Stephen J. Gould's collection of essays *An Urchin in the Storm* ranges from topics in geology and evolution to portraits of biologists. In his review (330, 280; 1987), Alan Brafield likened Gould to "his avian namesake the jay, colourful and garrulous". Price is £5.99. Richard P. Feynman's famous lectures on quantum electrodynamics, *QED*, were first reviewed in 1986 (320, 661) by P. Waloschek, who recommended that "anyone with a curiosity about physics today should buy it", price £4.99. Catherine Caufield's book on the "chronicles of the radiation age", *Multiple Exposures*, is published at £5.99. John Dunster in his review (341, 30; 1989) called the book gripping, but biased. But, he said, "books without bias make dull reading". Ian Stewart's explanation for the layman of the mathematics of chaos, *Does God Play Dice?*, was described by Robert M. May in his review (338, 686; 1989) as "engaging, accurate and accessible to the uninitiated". Price is £6.99.

■ Two new paperbacks from Oxford are *The Unheeded Cry*, by Bernard Rollin, and Martin Gardner's *Whys and Wherefores*. The former is subtitled *Animal Consciousness, Animal Pain and Science*, and is part of Oxford's Studies in Bioethics series, price is £6.95. The second book is the latest collection of Gardner's essays and reviews, subjects including the puzzles in Joyce's *Ulysses*, the fantasies of Ray Bradbury, Arthur Clarke *et al.*, and computer programs capable of discovering scientific laws. Price is £5.99. □

Little Hans, clever Hans and big Hans

Steve Blinkhorn

Rebel with a Cause: An Autobiography. By Hans Eysenck. Allen: 1990. Pp. 310. £14.95.

AS AN antidote to falling ratings, I propose the following plot for a television mini-series: handsome son of glamorous film star and excellent tennis player — no mean boxer — abandons promising career as physicist. Despite father's influence with Goering, flees Nazi Germany for England, escaping arrest by the SS and the concentration camps by seconds. Returns to Germany to smuggle out family jewels under nose of Gestapo. By chance, enrolls to study psychology; marries; completes PhD in wartime Britain, financed by wife.

War work in emergency psychiatric hospital; develops interest in hypnosis as a blind for introducing a bed into research rooms — real purpose, seduction of nurses. Cures case of impotence with gift of copy of *Fanny Hill*.

Professional advancement leads to establishment of own research department; political ideals bring many left-wing friends. Falls in love with glamorous daughter of famous musician; deserts wife and son to set up house with and later marry her, and lives happily ever after.

Writes popular books to supplement low academic salary, while inventing clinical psychology, behaviour therapy, personality theory and so on, at the same time dealing devastating intellectual blows to psychoanalysis, psychotherapy and all manner of pseudo-scientific superstition. All this while working only seven hours per day, two spent playing tennis.

Increasing fame leads to controversial stands on inheritance, race, intelligence, smoking, extra-sensory perception, astrology. Meanwhile, research department flourishes, standing in citation index grows, award of DSc.

The years pass; hair turns grey, but age cannot mellow him. Retirement brings yet more controversy, more books, more mould-breaking theories. Series ends as hero strides onto tennis court, still tall, still handsome, but now grave and distinguished; memories dance before his eyes, of battles fought and won, of students going on to found their own departments, and of the first girlfriend, back home in Germany, the *Hitlermädchen* he had to leave behind.

This is the story of Hans Eysenck,

simplified a little, dramatized a touch, but embroidered not at all, as he tells it. Of course the emphasis is a little different, but insofar as there is a personal story to tell, this is it.



“... with just a hint of the outrageous and ridiculous”.

A frequent reason for paying heed to the biography or autobiography of members of the intelligentsia is the expectation that their life histories will illuminate the development of their thinking. Psychologists may possibly be necessary exceptions to this principle. Certainly, Eysenck's line of attack is more to indicate how the theories he has developed or advanced shed light on the course of his life, which is a less than engrossing approach given the cool acceptance of the dictates of extramental forces that his approach to the psychology of learning and personality seems to indicate.

Throughout, there is a sense that events occur in obedience to the joint dictates of necessity and chance, in which motives, goals, values and desires, if they have a place at all, are little stressed and less regarded. Which, in terms of the proposed mini-series, leaves this reader a little bemused. The story is all plot mechanics and no character development, like a cross between *The Thirty-Nine Steps* and *National Lampoon*. The man is not central: his scientific pursuits are. Certainly those with any degree of personal acquaintance with our hero will readily recognize the self-portrait (and perhaps others will find it surprising: here is a man

whose public image is at variance with his personal style).

But the question does arise as to whether the self-portrait is any less a misreading of character than the media caricature of a superannuated *enfant terrible* — a caricature echoed in the book's title. Psychoanalytically inclined readers might well look to the suggestion of lack of parental attention in childhood, the break with roots in Germany, the insecure position of an enemy alien in wartime England, as elements in a dynamic with great explanatory power.

But if there is an element which does seem understressed in Eysenck's account of his progress through life, it is the advantage of an outsider sufficiently at home in his adopted culture to operate effectively without the baggage of oversensitivity to subtle cultural controls.

There can scarcely be a scientist who has managed simultaneously to develop a successful department, become an effective popularizer of his subject and be the centre of popular controversy (with just a hint of the outrageous and ridiculous) to the extent that Eysenck has. His *penchant* for pressing an argument just one inference too far is simultaneously headline-grabbing and irritating. I wonder too whether all concerned will feel happy

with his attributions of priority in the various fields of theory and research with which he has been concerned. As he himself puts it, British psychology in the 1940s was essentially feudal. The resonances of disputes now 30 years old or more can still be detected by those with ears well tuned to ancient echoes. Yet one never has the sense that here is a vain man, even when there are touches of vainglory in his intellectual campaigns.

But this is meant to be a book review, not a character analysis. And so it must be said that autobiography is not Eysenck's natural *métier*. Perhaps because it is the least intellectually pugilistic, least controversial and tendentious, and most narrative of his works, it is less than gripping stuff. As a matter of record, it would have been a pity had it been left unwritten. But as an idiographic account of one man's life, which it explicitly sets out to be, it is disappointing. What is more, “idiographic” is misspelt virtually throughout. A slip that might have interested Professor Freud? □

Steve Blinkhorn is Managing Director of Psychometric Research & Development Ltd, Brewmaster House, The Maltings, St Albans AL1 3HT, UK.

Seaside story

J. H. Steele

Biological Oceanography: An Early History 1870–1960. By Eric L. Mills. Cornell University Press: 1989. Pp. 378. \$47.95. Distributed in the United Kingdom by Trevor Brown, £26.

I ENJOYED reading this book. It is clearly written and is based on thorough research so that a great deal of historical information is conveyed effectively. But oceanographers, like most other scientists, are not particularly interested in the history of science, even if it concerns their own field. Are there other reasons, apart from curiosity, for reading this book? I think there are.

First, the subject matter. The history of biological oceanography is restricted here to developments in the study of plankton dynamics. The author regards basic productivity as the core of this field, but even within this category the work of some well-known oceanographers, such as Marshall and Orr, Steeman Nielsen and John Ryther, are mentioned only briefly or not at all. However, this allows Mills to focus on his main theme, the rise and fall of centres of excellence, especially those at Kiel University from 1890 to 1920, and the Marine Biological Laboratory in Plymouth from 1925 to 1940. Both of these centres had intellectual leaders — Hensen and Brandt in Germany and Harvey and Brandt in England — and each group involved several other active researchers. (It is relevant to note that none of these people had graduate training in his final field of research: Victor Hensen's initial research, in about 1860, was on glycogen and epilepsy; and W. H. Harvey took a degree in chemistry and worked in the family paint

business before going to Plymouth in 1921.)

Some members of these groups are remembered for their individual contributions but others are not. Among the latter group, I was fascinated to read about Alexander Nathanson who, between 1906 and 1910, defined the significance of vertical mixing between the upper layers of the ocean in the control of plankton production. According to Mills, Nathanson played a crucial part in the development of the Kiel group and his ideas foreshadow present interests. Yet I — and several of my colleagues in this field — did not know his name or his work. Similarly, the measurements by Penelope Jenkin at Plymouth in the early 1930s established the close relation of photosynthesis to light at different depths in the sea. I doubt if her work is referenced or even remembered now. Recounting such details, Mills makes the point that during their most productive periods these groups were more than an aggregation of talented individuals. They were brought together by the need to collect and interpret diverse kinds of data in ocean physics and chemistry as well as in plankton biology. Mills's analysis indicates that these centres were successful when the collective capabilities and facilities matched the nascent concepts.

The last main actor in Mills's story is the American, Gordon Riley, whose significant work on modelling plankton dynamics was published between 1946 and 1951. Riley worked at Yale University and at the Woods Hole Oceanographic Institution — and later in Nova Scotia. Mills calls him "a loner". He did not establish or participate in a centre in the Kiel or Plymouth tradition. His influence was predominantly through his publications. Mills considers that Riley's work "had surprisingly little impact outside oceanography and is often still unappreci-

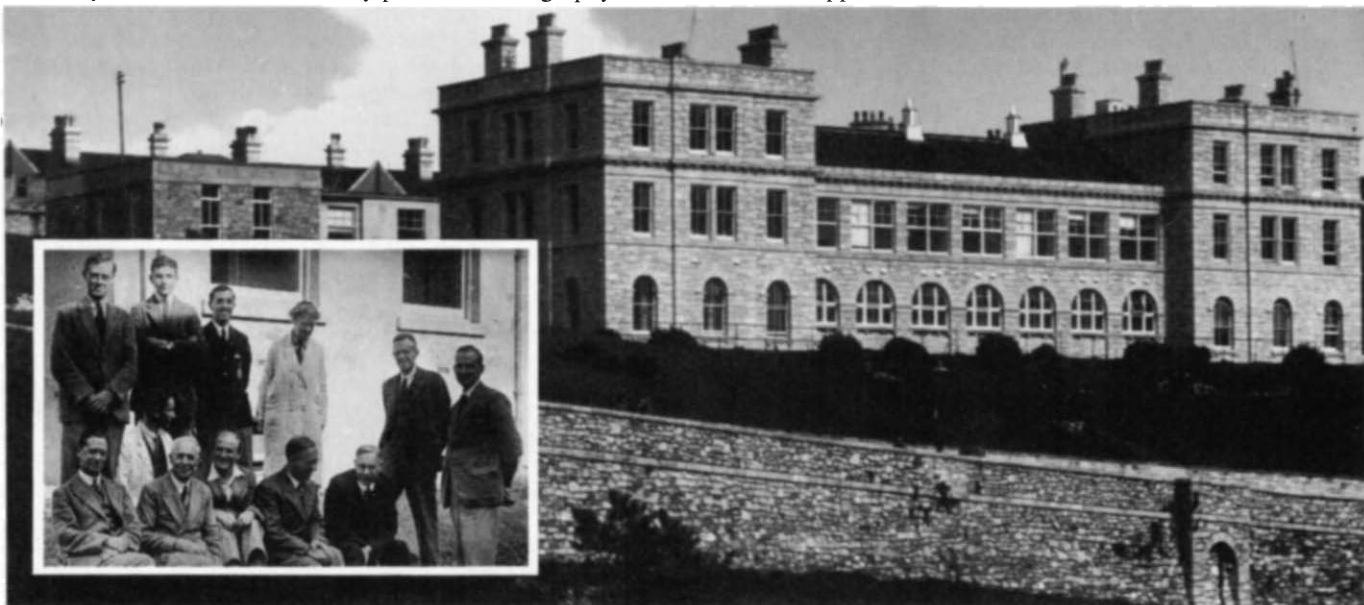
ated within it".

Personally, I place a high value on Riley's work, but Mills's implication is that the social structuring of research groups is a critical element in bringing to fruition new ideas in fields such as oceanography which require diverse technical facilities and close collaborations in field work. The significant impact of a later group established by John Strickland at the Scripps Institution in California shows that this approach of bringing together active researchers in an institutional setting was compatible with the US system and was still appropriate in the period 1960–70 (which is outside the time limits of this book).

In the past decade, new and diverse technologies such as satellites and electronic mail have partially removed the dependence of oceanographers on ships and daily face-to-face interactions. There has also been the emergence of a wider appreciation of global problems encompassing research across the land–sea–air sectors as well as across disciplines. Are the earlier institutional patterns from Kiel, Plymouth and Scripps appropriate to these more recent developments? We are seeing the creation of groups with common scientific interests recruited from several institutions, often in different countries. But the problem of matching scientific concepts to technical abilities still has to work within the limits of human relations. We are in the process of discovering these old constraints in a new context. The rise — and the fall — of previous groups can have a more than historical interest, and Eric Mills's book is relevant to these aspects of carrying out research effectively. □

J. H. Steele is at the Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA.

Marine Biological Association



The Marine Biological Association building in the 1930s. Inset, W. H. Harvey (sat on the left) and colleagues, 1933.

What the doctor ordered

John Galloway

Historical Perspectives on the Role of the MRC. Edited by Joan Austoker and Linda Bryder. Oxford University Press: 1989. Pp.259. £30, \$49.95.

A RESEARCH organization is judged — or should be judged — by what it is doing, not by what it has done. The scientific past tends to be insignificant compared with its present: "science swamps its own history", as Barry Barnes says in his book *About Science* (Blackwell, 1985).

Nevertheless, historians (and others) do write the histories of research institutions, some of which clearly like to have their histories, or at least authorized versions of them, written up. The UK Medical Research Council's 'authorized' version was written by Sir Landsborough Thomson, retired second secretary of the council, and published in 1973. What we have in the present volume is not an alternative complete history of the MRC but rather ten essays about different aspects of its work, written with one exception by professional historians. The exception is Sir Christopher Booth, last director of the Clinical Research Centre who is, I think it fair to say, a distinguished *amateur* historian.

Sir Landsborough's history, written as it was by a man who spent most of his working life with the MRC, is presumably the nearest thing an institution can have to an autobiography. It is this rather than the fact that it was authorized that is worth remembering — the essays in the present collection are in a sense no less authorized; they were written with considerable access to the MRC's archives, or at least such as were left of them. (One result of Sir Landsborough's version, if I remember correctly from my days at MRC, was that many archives were destroyed, an action justified apparently by the belief that whatever history they contained had been transferred intact to his book.) The writers also had access to members of MRC staff past and present. What they could bring to the task, which Sir Landsborough, of necessity, could not do, is an alien viewpoint — a searching, intrusive look from outside.

The viewpoint is alien in more senses than one. The research described in these essays was almost without exception done by men, as was the work of the committees monitoring its progress and taking decisions about it. But five out of the eight essayists are women. There is a lesson here. For those who believe — and who can blame them? — that medical research in Britain began round about

1950 with the startling insights provided by molecular biology (Nobel prizewinners Crick, Watson, Sanger, Perutz and Kendrew all worked in MRC laboratories), these essays may come as a surprise. To begin with, almost all the research they relate took place before then. In scientific terms this is not merely history, but largely ancient history. And second, the research is very applied: tuberculosis, tropical medicine, industrial health, public health (in two essays),

applied new research findings and new statistical methods to what were perceived to be the pressing medical and social problems of the day. Indeed, it is argued persuasively by Linda Bryder that it was explicitly the problem of tuberculosis — which assumed enormous proportions in the early years of this century — that brought the MRC into being. Concern that Britain was lagging behind Germany in the treatment of tuberculosis and so in 'national efficiency' led to this disease



Park Crescent — home of Britain's Medical Research Council since 1961, built to conform with the Georgian façade of the other buildings surrounding it.

insulin used as a model for relations between the MRC and the pharmaceutical industry, experimental radiobiology and clinical research. There is also an essay about the National Institute for Medical Research and a biographical essay on Sir Walter Fletcher, the council's first secretary.

What is striking is the extent to which the council created knowledge and

being the only one for which free institutional treatment was provided under the terms of the 1911 National Insurance Act. An apparently late appreciation that science might permit the state-provided funds to be spent to the best advantage led to the inclusion of a clause in the act setting aside one penny per person insured (amounting to £56,000 a year) for research. It seems clear that this research

element was regarded as a minor point of drafting, although it turned out to be a prescient one. Whether it was intended that research be limited to tuberculosis is not now clear, although Bryder thinks it was. Whether it was or not, pressure was immediately applied to enlarge its scope.

Today, the notion of objectively assessing new forms of treatment to see whether they achieve anything or nothing through controlled clinical trials is commonplace, if not always uncontroversial. In Britain this can be traced back to the MRC's therapeutic trials committee which, among other things, tested calceferol, testosterone and the sulphanilamides. Interestingly, the committee had no statistician: even though R. A. Fisher pioneered the idea of randomization in agricultural experiments and worked with the MRC in several research arenas, he apparently had no links with it. Possibly because of this, the idea of the randomized controlled clinical trial had to wait until 1946 when Bradford Hill used such a design to test streptomycin as a treatment for tuberculosis. The results left little doubt about the benefits of the drug. But much more important, they ushered in a new era of medicine.

Just as innovative but, with the benefit of hindsight, rather more suspect results had been the work of the committee of quantitative problems in human nutrition, first convened in 1922. The committee analysed working-class spending on food and found that the working classes were "ignorant and inefficient". In fact, as Celia Petty points out in her essay on the relationships between research and public health, the data the committee collected "actually demonstrates that working class wives already made extremely efficient use of inadequate resources and had a better grasp of economic realities than their academic critics".

Lessons can be learned from this particular piece of history. The first is that those doing the study in the years between the two world wars presumably found what they expected to find (as Thomas Kuhn had told us they would). What was expected here was the view of one class of society about another. Nevertheless, the work was a novel attempt to "apply modern scientific methods to improve the well-being of an economically disadvantaged group". However, the fact that the wrong conclusion was drawn probably had far-reaching consequences as it supported the idea that the lower paid did not need more money, but more 'education' (this view seems to have resurfaced recently). In fact, they were simply underfed.

How difficult it is to apply science successfully to social problems. The irony revealed by Petty's essay is that the nutritional research did not directly alleviate the social problems which were its motivation, but rather led to better understand-

ing of physiology and biochemistry. In the short term, it was science not society that benefited. The late Derek de Solla Price made much the same general point about science and technology. I think that if there is one thing that the research community should heed in this day and age, it is that particular lesson.

Our research institutions play large, though for most of us largely unknown, parts in our lives. We owe it to ourselves, as well as the institutions themselves, to try to understand them. There is a school of thought that understanding an institution or an activity requires knowing its history. This is not a notion to which scientists themselves often seem to subscribe, perhaps with good reason as I suggested above. It seems to me, however, that we fail at our peril to understand science and particularly how research can be harnessed to social change through political will and legislation or 'market forces'. We are not short of accounts of science as threads in the history of ideas. But this is the age of science by government, and anything that provides even a glimpse of how science drives social engineering through the interlocking committees that are the cogs of government machinery is profoundly to be welcomed.

These are essays academic in tone, and this is their strength and their weakness. They are well-documented and well-argued, but they are also rather circumscribed and tend to leave the reader slightly dissatisfied. Bryder's judgement on the MRC's contribution to the decline of tuberculosis is hardly rhadamanthine in its severity. Nevertheless, she does say that the reduction of the disease occurred largely independently of "medical intervention". But to find out what did cause its reduction, she refers the reader to two other books (neither of which I could find in a library). Booth's history of clinical research stops at the point the story starts to get interesting — the opening of the Clinical Research Centre. It is that institute's closure 20 years on that has been arousing passions recently. We really do need to know why this enterprise failed.

A final word of warning. This is administrative history — it depends for its raw material on official documentation. One does not have to sit through many committees, research or otherwise, to appreciate that what is written in the minutes bears only a limited resemblance to what has been said. I sometimes think historians should be more like the psychiatrists who believe nothing their patients tell them rather than like those who believe everything. □

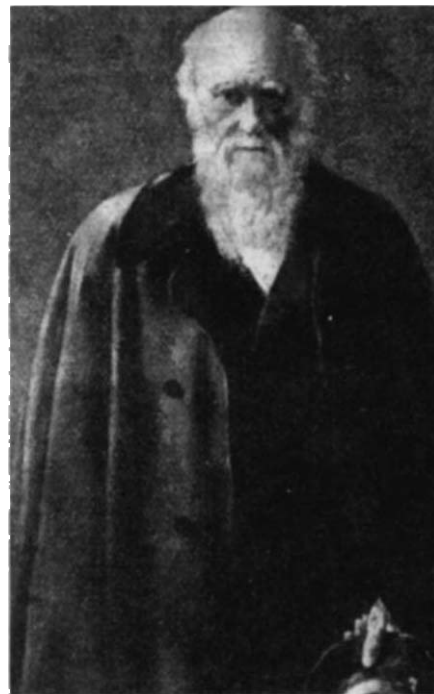
John Galloway is at the Cancer Research Campaign, 2 Carlton House Terrace, London SW1Y 5AR, UK. The views expressed here are his own.

Darwin's legacy

Philip Kitcher

The Metaphysics of Evolution. By David L. Hull. *State University of New York Press: 1989. Pp.331. Hbk \$73.50, pbk \$24.95.*

MANY practicing scientists have a stereotype of the philosopher of science which can be condensed into three characteristics: ignorant, interfering and insular. That stereotype derives from the suspicion that the philosopher of science is primarily in the business of telling the practitioners what they ought to be doing, and that this unwanted advice is given



Darwin — not preferring the status quo.

from an olympian perspective.

Work in the philosophy of physics since the 1940s and in the philosophy of psychology since the 1960s has belied the stereotype, for philosophers in these subdisciplines have demonstrated their ability to draw on the substance of the sciences and to contribute to the clarification of theoretical disputes. During the past two decades, however, biology — particularly evolutionary biology — has been the most exciting area for exchange of ideas between philosophers and scientists. David Hull is one of those most responsible for this dialogue, and many of his contributions to it are brought together in this volume.

Hull is a vigorous campaigner against what he calls "intellectual territoriality", and his essays range over issues in metaphysics, history of science, general history, systematics, evolutionary theory and sociology. He offers insights into the reception of darwinism, and uses a study

of reactions to *The Origin of Species* to argue cogently against the popular idea that older scientists prefer the *status quo* whereas their younger colleagues are more innovative. He provides illuminating discussions of trends within contemporary systematics, taking those trends to illustrate some general processes that occur in the growth of science. This collection also includes an address "on human nature", celebrating the variability of species in general and our own species in particular; three defences of the thesis that species are individuals (of which Hull and Michael Ghiselin are the principal advocates); some essays on sociobiology which hew a characteristically ecumenical line; and some articles about the practice of the history of science.

Although his studies take him into numerous specialized fields, Hull's essays reveal a thorough command of his material. His writing is direct and clear, and he presents interesting and original ideas on a variety of topics. In some instances, he gives us refreshing doses of common sense. "In Defense of Presentism", for example, is a penetrating critique of some fashionable suggestions about the history of science. Recognizing the dangers of distorting the thought of earlier investigators, historians of science sometimes inveigh against any forms of "presentism" or "Whiggism", insisting that historical studies must eschew any categories that would not have been accepted by the historical actors. Hull argues forcefully that this position is both unmotivated and absurd. We can appreciate the danger that our current beliefs might distort historical understanding without yearning for histories that uncritically accept the beliefs of earlier generations. As Hull points out, the historian "has every right to believe that virgin births were no more common two thousand years ago than they are today" (page 217).

Not all of Hull's essays are equally successful. His central argument for the view that species are individuals rather than classes seems to depend on misunderstandings of the notion of a class. Classes (sets) are atemporal entities that do not change over time. How then, Hull asks rhetorically, could a species be a class? The answer is simple: we think of the evolution of species in terms of the differences in the distributions of properties at different temporal cross-sections. The atemporality of classes generates no sound argument for Hull's favoured conclusion, and, in consequence, no rationale for the views about the character of macroevolution that he and others have defended by appealing simply to metaphysics. Many practicing biologists, as well as philosophers, have found inspiration in the idea of the individuality of species. Other biologists and philosophers (including myself) see it as much ado

about very little.

The last two essays in the volume, reflections on human sociobiology and on the sociobiology debate, also seem less well argued. When first published, they struck an admirable tone of mediation: criticisms of sociobiology were to be elaborated, but they should not be allowed to throttle a research programme in its infancy. More than a decade later, the dialectic has unfolded and, in the absence of some rethinking and rewriting, Hull's irenicism is bound to seem flabby and dated.

In the end, however, it is churlish to

complain about a collection offering so many provocative and well-defended ideas about so many topics. Biologists, historians, sociologists and philosophers should be glad to have this convenient synopsis of the ideas of an influential thinker. Hull has done more than his bit to fashion a new stereotype of the philosopher of science, founded on a different trinity of traits: informed, incisive and interesting. □

Philip Kitcher is in the Department of Philosophy at the University of California, San Diego, California 92093, USA.

Watch the birdie

John R. Krebs

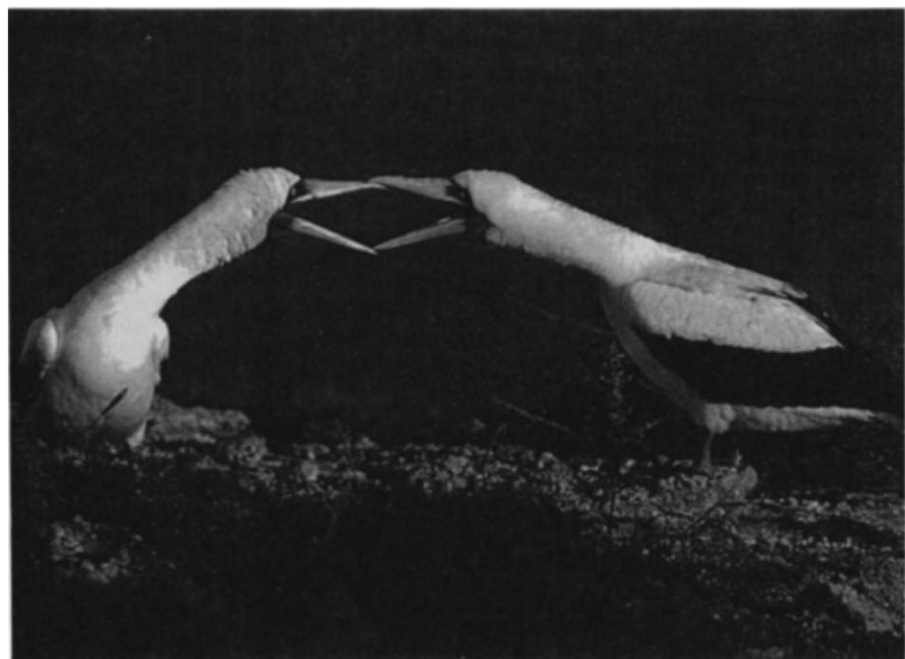
Ornithology. By Frank Gill. Freeman: 1990. Pp. 660. £29.95, \$49.95.

BIRDS impinge on our lives in many and diverse ways. Some people see them as a symbol of the beauty of nature, to be observed and conserved in their natural habitats (the Royal Society for Protection of Birds, I am told, has more paid-up members than either of the main political parties in Britain). Others enjoy their company as talking or singing household pets; as agile quarry viewed down the barrel of a shotgun; or as a convenient snack purveyed by Colonel Saunders. This last group will find little to interest

Kingdom), will find it contains much to inform and enjoy.

Roughly three-quarters of the book is about ecology and behaviour of birds, the rest being concerned with morphology, physiology and evolution. To test the book as a source of ornithological knowledge for students, I looked up the answers to four questions which I have been asked by students in the last year or so. I also compared Gill's answers with those given by the main rival texts: J. Faaborg's *Ornithology: an Ecological Approach* (Prentice Hall, 1988), and the fourth edition of J. C. Welty's and L. Baptista's *The Life of Birds* (Saunders College Publications, 1988).

Here are the questions and answers. (1) How do birds produce sounds? Greenewalt's (1968) theory of the internal tympaniform membranes as the sound source is



Nesting Boobies — a love affair in the Galapagos.

them in Gill's book, but others, as well as the undergraduates for whom it is primarily written (ornithology is still taught as a separate subject in North America although not in the United

simply but accurately summarized by Gill, although Nowicki's important (1987) result showing the effect of helium on the harmonic structure of sounds is not mentioned. Neither of the other two texts get

Oxford Scientific Films

as far as Greenewalt. (2) Why do bird sounds in different habitats have different characteristic time and frequency patterns? Gill presents the two main current hypotheses: sounds are designed by natural selection to overcome habitat-specific attenuation (Morton) or degradation (Wiley). The other two books mention only Morton's hypothesis. (3) Why do birds perform elaborate courtship displays? The accounts of both Welty and Baptista and of Faaborg start from the traditional view that courtship serves to "synchronise sexual arousal" or "maintain the pair bond". Gill, on the other hand, uses a more modern conceptual framework within which males and females are seen to have conflicting interests, so that much of courtship has to do with assessment of partner quality and fidelity. (4) Why do some gamebird populations show regular cyclic fluctuations? The red grouse is taken as a case study by Gill, and he summarizes the Watson-Moss view that cycles are driven by behavioural changes. The alternative view of Hudson, that the gut parasite *Trichostrongylus tenuis* causes the cycles, is not mentioned. But the other two books do not give worked examples and only make very general statements about possible causes of cycles such as food, predators and behaviour.

My test shows that Gill's book, although superficial in places, is more up-to-date and incisive than its rivals. It is also most attractively produced. I would recommend it as the preferred text for an ornithology class, unless it was essential to have the sections in the other two texts on "Birds and Man" — dealing with applied problems not covered as a separate issue by Gill.

Let me end by telling an historical anecdote. On page 9 of his introduction, Gill asserts that "Albert Szent-Georgy (*sic*) won a Nobel Prize for the elucidation of the Krebs cycle from studies of pigeon breast muscle". The crucial discovery for the cycle (conversion of oxaloacetate and pyruvate (via acetyl CoA) into citrate) was reported by H. A. Krebs (in a paper famously rejected by *Nature* on 14 June 1937). Szent-Györgyi discovered that several 4-carbon dicarboxylic acids are readily oxidized in suspensions of minced pigeon muscle and that these dicarboxylic acids can act 'catalytically' to accelerate the combustion of substances in the muscle suspension.

I remember once, as an undergraduate, hearing Szent-Györgyi in a lecture at Woods Hole claiming that the citric-acid cycle should have been named after him, not Krebs. I wonder if Frank Gill was in the same audience? □

John R. Krebs is in the Edward Grey Institute of Field Ornithology, Department of Zoology, Oxford University, Oxford OX1 3PS, UK.

The science of queens

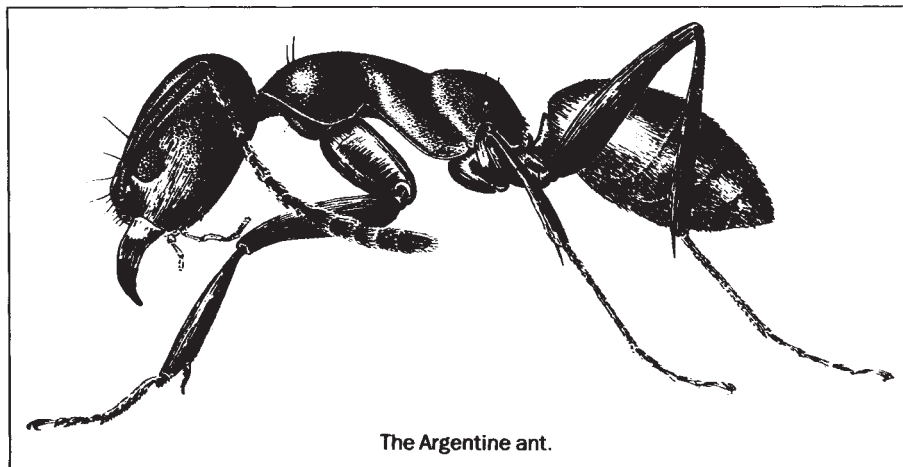
Donald H. Feener Jr, Diane W. Davidson and Jon Seger

The Ants. By Bert Hölldobler and Edward O. Wilson. *The Belknap Press of Harvard University Press*: 1990. Pp.732. \$65. Distributed in the United Kingdom by Springer-Verlag, £70.50.

INDIVIDUALLY, ants are small and unremarkable; collectively, they rule the terrestrial world. They live almost every-

field guide, and we hope that the authors will do this. (At more than 2.5 kilograms, *The Ants* is not portable.) Other chapters include extensive tables summarizing biogeographical distributions, regional faunistic studies, modes of colony formation, pheromone chemistry, and many other subjects. There is a glossary of several hundred terms, and some chapters have what amount to mini-glossaries of their own. The bibliography (65 pages) contains 3,000 references. We doubt that any major taxon has ever received such a thorough and synthetic review.

With uninhibited enthusiasm (and flashes of humour), Hölldobler and Wilson begin *The Ants* with brief accounts of



The Argentine ant.

where and eat almost everything. They make up only 0.1–1 per cent of all animal species but 10–25 per cent of the animal biomass in many places. Ants have pushed insect sociality to its limits in several different respects, and this is in large part the key to their ecological success. Yet they are hardly mentioned in many ecology textbooks, and "not one biologist in a hundred can describe the life cycle of any species". This state of relative neglect should soon end, because Bert Hölldobler and Edward Wilson have put most of myrmecology between covers.

The Ants is a stunningly attractive volume that belongs as much on the coffee table as it does on the lab bench. It contains nearly 1,000 illustrations, including 24 spectacular colour plates that vividly portray the lives of various species. (How many children will slowly turn the pages of this book, and years later take up the study of ants?) The 20 chapters are organized thematically, and they are written in a clear, accessible and engaging style. The chapter on classification is a book within a book; it includes anatomical drawings and definitions, a genus-level taxonomy, keys to the 11 subfamilies and the 292 currently recognized genera (50 pages), and habitus drawings of every genus (60 pages). With this extraordinary tool, even a novice will be able to identify any ant to genus; the chapter could be made into a splendid

the ecological and scientific importance of these and other social insects. Insiders will be thrilled, and even the most sceptical should be convinced that "the modern insect fauna has become predominantly social". In chapter 2, on classification and origins, the authors estimate that only about half of the world's ant species have been formally described. The vast majority of unknown species are confined to rapidly disappearing tropical rain forests — Wilson recently collected 42 species in 26 genera from a single tree in the Peruvian Amazon, for example. Although the taxonomy of ants is in better shape than that of many other insect groups, many phylogenetic relationships remain unclear, including those of the 11 subfamilies. These gaps in our understanding should soon yield to rigorous phylogenetic analyses applied to both morphological and molecular data.

Chapters 3 to 9 review what is known about colony life cycles and interactions among colony members; the topics include altruism and the origin of workers; differentiation and division of labour among worker castes; colony odor and kin recognition; communication; and colony homeostasis. Chapter 10, on foraging strategies, territory and population regulation, illustrates the authors' focus on general biological principles. In arguing for the importance of competition in ant communities, they draw on more different

kinds of evidence than are presented in any recent general ecology text; here the reader will learn about overt aggression, overdispersion of colonies, overdispersion and complementarity of traits, exclusion experiments, packing experiments, natural and experimental introductions, and several instances in which introduced species have subsequently evolved reduced levels of aggression.

In the next four chapters, the authors consider the symbiotic relationships of ants with one another, with other arthropods, and with plants. Again, general principles are stressed, but not at the expense of natural history. Then follow five chapters on the evolution and ecology of the most distinctive ant lifestyles: the specialized predators, the army ants, the fungus growers, the seed harvesters and the weaver ants. A final short chapter describes various techniques for collecting ants and for culturing them in the laboratory.

Only Hölldobler and Wilson could have written such a comprehensive and integrated treatment of ant biology. It represents a herculean labour of love, and it sets a new standard for synthetic works on major taxa. Every expert will find something to complain about; each of us has a list of quibbles concerning overlooked literature, misplaced emphasis or questionable interpretations of evidence, in those particular areas we happen to master. But such flaws are inevitable in any book that attempts to organize a large and complex field, and they do not detract from its value in any important way. Like a social-insect colony, *The Ants* is much more than the sum of its parts. (Incidentally, why entire ant colonies behave in ways that are far more precise and reliable than the behaviours of individual workers is explained in chapter 9.)

The Ants can be viewed as a sequel to Wilson's *The Insect Societies*. Published nearly 20 years ago, that book was a similarly ambitious synthesis of natural history, experimental ecology, physiology and evolutionary theory. It convinced several generations of graduate students that the termites and the social Hymenoptera (ants, bees and wasps) are inherently fascinating organisms, and that they can be used very effectively to address fundamental issues in behaviour, ecology and evolution. Despite its age, *The Insect Societies* is still in print and still invaluable; our well-worn copies are always close at hand. Like its predecessor, *The Ants* will undoubtedly remain in active service for decades, guiding both tourists and seasoned travellers through a strange and wonderful world. □

Donald H. Feener Jr, Diane W. Davidson and Jon Seger are in the Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA.

Guardian of the gates

Steven Dickman

The Cuckoo's Egg. By Clifford Stoll. Bodley Head: 1990. Pp. 326. £12.95. Published in the United States by Doubleday: 1989. \$19.95, and in Germany by Fischer: 1990. DM29.80.

PEOPLE who break into computers no longer do it just for fun. What used to be a diversion for lonely teenage savants has now become big business for criminals and spies.

A case in point was when West German 'hackers' — as these computer burglars have come to be called — in 1986 and 1987 broke into a number of United States



Stoll — ebullient charm.

military and academic computers. Gaining access through open academic networks, the hackers penetrated supposedly secure computers at US defence contractors and at military bases from Alabama to West Germany to Okinawa.

Although the hackers may have begun with innocent motivations, they quickly discovered that computer crime does pay, especially when military secrets are at stake. Until they were arrested in 1989, the hackers 'earned' DM90,000 (about \$53,000) for delivering information to the KGB.

In the end, not that much damage was done to US national security; none of the information the hackers stole turned out to be classified. But the psychological damage was immense. For most computer users, military or otherwise, the thought of having intruders tromping through data files or changing the operating system are enough to bring on a cold sweat.

Fortunately, the individual user has an ally in fighting the battle for computer security — Clifford Stoll, a mopheaded astronomer-turned-computer-sleuth with a weakness for chocolate-chip cookies. Just a week after starting a job as a computer jack-of-all-trades at Lawrence Berkeley Laboratories in California (where, Stoll says, they "recycle used astronomers"), he was confronted with an accounting problem: the internal records showed an error of 75 cents. He thought it might take an afternoon to locate the source of the error. But when it became clear what was at stake, Stoll stayed up at night for two years, tracking a professional intruder nine time zones away who was rifling through dozens of computer systems.

Stoll narrates the tale of the chase in his fast-paced book *The Cuckoo's Egg*, from the first puzzling day when the computer's books did not balance until the triumph more than two years later when a hacker (and later his accomplices) was arrested in Hannover, West Germany, and charged with espionage. Stoll's editors must have faced a tough decision when they read his unique jargon-laced prose: whether to clean up the hippy, computer and California slang in the interests of clarity, or to leave it to the reader to make sense of it.

They chose the latter, allowing Stoll's ebullient personality to shine through. The British publishers have left the vocabulary alone, a decision that might leave non-technical (or non-US) readers justifiably perplexed by words like 'grinch' and 'kludge'. But despite a few rough spots, even someone completely 'computer-illiterate' will be able to follow the chase.

Much of the book's charm lies in Stoll's apparently unembellished descriptions of his extracurricular activities as an average hippy in Berkeley — knitting quilts, going to costume parties, bicycling in the rain and drying his sneakers in the microwave oven. In between the anecdotes and the chocolate-chip cookie recipes, Stoll describes in painstaking detail how he made hacker-hunting a science, complete with lab notebook, scientific meetings and publications. His detailed records became valuable documents that were later used to reconstruct the break-ins before a West German court, where the hackers were eventually convicted of espionage.

Despite his continuing reluctance to identify himself as anything but an astronomer, Stoll belongs to a new — and necessary — breed of public servant: the hacker tracker. In his book, he reveals how little anyone — users, network managers, even intelligence agencies, had thought about security. Because of his efforts and the publicity they attracted, Stoll himself has already brought about some improvements in the system.

Military networks, recognizing that the

stakes have been raised, are responding to the threat by installing new techniques such as a form of caller identification as well as simply by cutting themselves off from the outside world — the ultimate in computer security. Some businesses and scientific institutions are following suit.

But academic computer systems, especially those which offer on-line access to remote users, still face some difficult

decisions. In some operating systems, increasing security necessarily means restricting access, which defeats the purpose of open networks. Clifford Stoll and his fellow guardians of the gates will have their hands full in the coming years in finding a happy medium. □

Steven Dickman is Munich Correspondent of Nature.

Packing quite a punch

Simon Lavington

ICL: A Business and Technical History. By Martin Campbell-Kelly. *Oxford University Press: 1990. Pp. 409. £30, \$30.*

THE British computer industry, like the British car industry, is now firmly established as fair game for comment in saloon bars across the land. There are other similarities: computers, like cars, affect us all, and, as both products struggle to become more user-friendly, familiarity breeds contempt. Martin Campbell-Kelly's scholarly history of ICL (International Computers Ltd) pulls us up sharply. ICL is, for all practical purposes, the United Kingdom's only major computer company. It deserves to be taken seriously.

ICL has its origins in two punched-card office equipment firms: the British Tabulating Machine Company and the Accounting and Tabulating Machine Company (later Powers-Samas). Although the heyday of the punched-card machine came during the 1930s, this style of equipment saw service from the turn of the century through to the early 1960s. Indeed ICT, the successor company to BTM and the Acc and Tab, was still making 85 per cent of its money from the sale of punched-card equipment in 1960. It is therefore not surprising that the general-purpose computer does not come into the picture until well over half-way through this book. Computer buffs will want to skip the first 50 years of history and start at the point when the fondly remembered British computer manufacturers such as Elliott Brothers, English Electric and Ferranti became involved in the series of mergers with ICT that led to the formation of ICL.

This book is not, however, written

for computer buffs. It is an analysis of business opportunities in the evolving world of information technology. It is, above all, about market forces and the contrasting responses to these forces on both sides of the Atlantic. In this respect, it gives a series of case studies in technical and marketing strategy that should appeal to a wide audience. If the office-appliance industry was the barometer of the economy in the mid-1920s, its successor — the information-technology industry — will certainly be the economic barometer of the 1990s. Economic forecasters should read this book, heed its lessons and advise management accordingly.



1938 Powers-Samas sterling multiplying punch.

There is much that Campbell-Kelly could have said about the management of ICL. Instead, he contents himself with the gentlest of quotes from others. Thus Arthur Humphreys, who rose to become the company's managing director, is of the opinion that: "the British Tabulating Machine Co. was really run by English gentlemen, and I think the business came fourth after hunting, shooting and fishing". As for commercial success or setback, Campbell-Kelly lets the facts speak for themselves. It is a feature of his book that tables abound, charting the announcements, delivery dates, sales volume and

value of all manner of office and computing equipment from 1908 to 1985. These are always set in context, with comparative figures being presented for US and, latterly, continental European manufacturers. Through these figures we see the company's fortunes periodically rise and fall. To be fair to ICL, computer market trends have been notoriously difficult to predict. There is the oft-quoted remark of an undisputed computer expert, Professor Douglas Hartree, in conversation with 'the' Ferranti computer salesman in September 1951: "We have a computer here in Cambridge; there is one in Manchester and one at the NPL. I suppose there ought to be one in Scotland, but that's about all."

If 'scholarly' is the natural adjective to describe this book, what qualifications are demanded of its readers? They certainly need to be computer literate, not only to overcome the modest number of technical terms encountered here but also, more importantly, to appreciate the differences in power and scope of the various computing systems mentioned by Campbell-Kelly. This is especially true of ICL's developments since 1970, as the company struggled with the technical complexities of its "new range". The architecture of this new range is, in the words of Campbell-Kelly, "masterly and in advance of anything offered by any other manufacturer". We are not told why. Similarly, two very innovative ICL ideas — the distributed array processor and the content-addressable file store — are left dangling in the text in much the same way as the projects were left dangling for years within ICL before being presented in a realistic way to the market place. This book is thus technically tight-lipped.

For those with a taste for the technical, the book nevertheless contains some appetizing morsels. For example, there is the accounting machine invented by the Spaniard F. P. Campos in 1923. Containing 1,000 registers, it also offered Babbage-like possibilities for mathematical computation — if only it could have been made to operate reliably. There is also the important work done by the British Tabulating Machine Company for the 1939–45 war effort at Bletchley Park. Finally, there is the innovative work done by Ferranti and at Manchester University in connection with the Atlas computer which became operational in 1962. It is interesting to note that the ex-Ferranti West Gorton factory and Manchester University have again combined to provide much of the input to ICL's latest research project: a novel parallel computer called EDS, the European declarative system. □

Simon Lavington is in the Department of Computer Science, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK.

Getting it down on paper

Walter Gratzer

The Pencil. By Henry Petroski. Knopf: 1990. Pp.434. \$25.

"WHERE on earth is my teeny weeny penis?", demands Aldous Huxley's wagish edwardian pedant, Mr Beavis, to the inexpressible embarrassment of his son. "*Penecillus*", Beavis explains, "diminutive of *peniculus*, double diminutive of *penis*; which as you know originally meant a tail". For the primal pencil, as Henry Petroski also reveals, was a brush made from the tail hairs of domestic animals (a derivation that survives in the German, *Pinsel*, for a brush).

Etymology aside, the pencil appeals to Professor Petroski as one of those unregarded objects, like buttons or scissors, which nonetheless have a rich social and technological history. Pencils for many of us are numbered among the totems of childhood — pleasingly captured in a pair of Norman Rockwell pictures that Petroski reproduces. In one, "Back to School", a benign old codger behind the counter of a drugstore hands a Dixon Ticonderoga to an earnest little boy with a satchel. Venus and Koh-I-Noor are surely more names laden with nostalgia —

so familiar that *1066 and All That*, the epitome of English schoolboy humour, has it that Queen Victoria was presented at the Great Exhibition with a gigantic pencil called the Koh-I-Noor.

According to Petroski, a pencil was often known in the last century as a Faber. The French Caran d'Ache was named after the illustrator, Poiré, who took this as his pen name (so to speak) from the Russian *karandash*, a pencil. And how many people know that the familiar crazed green design on the Venus was the result of a faulty batch of lacquer — an effect judged to be distinctive enough to preserve?

Petroski is an engineer, the author of the admirable book *To Engineer is Human*. He has chosen the story of the pencil as an exemplar of technological evolution and its relation to economic circumstance and the desires of the consumer — largely, of course, before our own epoch of rampant technology, which has made invention the mother of necessity. Petroski's story begins with the

observation that a metal point could be used to make black marks on surfaces; this led to the search for more responsive materials, culminating in the discovery of the special properties of graphite, otherwise plumbago, black lead, *ochra nigra* or wadd, from the famous Borrowdale mines in Cumberland in northwest England. Soon, shafts of this material were encased in cedar wood, which is still used to make the best pencils.

The next great advance came when Conté, a French engineer, was enjoined to find a substitute for Borrowdale plumbago, which had become unavailable with the onset of the Napoleonic wars. The Conté process consisted in mixing, moulding and calcining cheap powdered graphite with clay and water. And so by degrees, with adjustments in the proportions of the components and in the procedures by which they were treated, and also the incorporation of new materials, such as waxes, it became possible to make pencils of reproducible quality and varying degrees of blackness (B) or hardness (H). In the United States, one of the first pencil manufacturers was the father of Henry David Thoreau, and the thriving cottage industry was continued by the sage of Walden Pond himself.

Petroski does not neglect the sociology or the economics of pencil making. With the expansion of demand came the search for new sources of suitable graphite. In the mid-nineteenth century, a French trader, Jean-Pierre Alibert, discovered a rich deposit in eastern Siberia, where with the assistance of the Russian government he set up a self-supporting colony of workers. Towards the end of the century, the pencil was replacing the slate in schools and, by 1870, 20 million pencils were being consumed every year in the United States. One reformer there noted that the written language contained one redundant (silent) letter in every seven, and so a seventh of all wood pulp (and graphite) production was wasted — a reflection to appeal to the modern ecologist, and reminiscent of the thrifty old Duke of Portland, who always left his 'i's undotted and 't's uncrossed to save ink.

Petroski has filled more than 400 pages with fact, fancy and cogitation about his subject. His treatise is stuffed with recondite knowledge, abundantly spiced with anecdote, and engagingly written. I confess to experiencing a slight sensation of surfeit by the time I arrived at the appendices, while conscious that I was in the presence of a notable contribution to technological history. The publishers have done their author proud: the book is most handsomely produced, the illustrations felicitous, the index a model and it is worth the money for the jacket design alone. □

Walter Gratzer is in the MRC Cell Biophysics Unit, Drury Lane, London WC2B 5RL, UK.



Mary Evans

Prized recipe — note taking in a cookery class, London c1900.

Into the atom

R. H. Dalitz

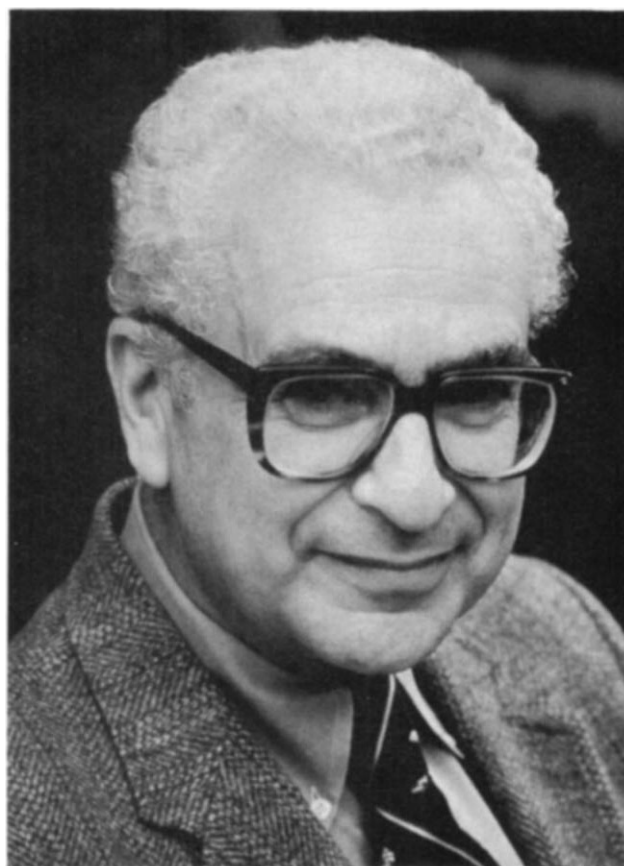
The Atomic Scientists. A Biographical History. By Henry A. Boorse, Lloyd Motz and Jefferson Hane Weaver. Wiley: 1989. Pp. 472. £19.65, \$27.95.

THE development of our knowledge and understanding of the particulate nature of the matter around us is traced by the authors of this book. Boorse, Motz and Weaver describe the field in terms of the lives and personalities of the main contributors to it, and provide substantial accounts of their research achievements.

The book is divided into sixteen chapters, the title of each indicating qualitatively the area of development covered. Each chapter has a number of sections, each with a specific title including the names of the scientists (usually one or two) who, in the authors' views, made the dominant contributions. Some scientists have their name in more than one of these section titles; for example, Rutherford's appears with "The Discovery of the Alpha and Beta Rays from Uranium", "The Transformation of the Elements", "The Alpha Particle and Helium", "The Nuclear Atom" and "Transformation of an Element", whereas Bethe's name appears with the inapt section title "High Energy Physics" in the chapter "Newer Developments in Atomic and Nuclear Theory" and with "Energy Production in Stars". A biographical sketch is given for each scientist named in a section title, often as long as two or three pages, generally in the first section whose title includes his name but sometimes spreading over several.

There are no diagrams nor displayed equations; the text includes a few mathematical expressions and several equations. Boorse, Motz and Weaver are brave men indeed to set out to explain, for example, the quantum mechanics of Dirac, Heisenberg and Schrödinger, or Maxwell's electrodynamics, or non-abelian gauge theories, to non-scientists, under these constraints. They have done very well overall, with some fairly skilful writing. I found only a few paragraphs where I did not at once recognize what was being described. How well a non-scientist would manage is not so clear; the text appears rather forbidding, being relieved only by the breaks between the longish paragraphs and by the small portraits of the scientists named in that section title. In general, the photographs used have not reproduced well, being much too dark; a few have obviously been touched up and at least one shows a distorted, atypical face. Many portraits are missing in the first third of the book, not surprisingly, and the reader feels their absence.

A number of the scientists included are not at all well-known — the names of Herapath, Waterston, Villard and van der Broek appear in section titles, to mention a few — about half of the entries without portraits are for these and other little-known scientists. The individuals chosen are sometimes surprising: Jensen is cited in "The Nuclear Shell [model]" section title, but not Maria Goeppert-Mayer, who shared their half of the Nobel prize in 1963; E. P. Wigner merits a place in this



Murray Gell-Mann — out in the cold.

company, but his name does not appear in the index, although it is mentioned in the text.

The outstanding omission in the section titles is the name of Murray Gell-Mann, the leading theoretician during all the decades of exploration and comprehension of the strange particles, the K-mesons and the hyperons, and who received the Nobel prize in 1970 for this work. The final section in the concluding chapter, "High Energy Physics", is entitled "Elementary Particles" but bears no name. Surely the name of Gell-Mann would have appeared most natural in that place? He introduced most of the concepts which are now taken for granted in elementary particle physics. Gell-Mann first introduced the classification system we use for these particles, and he first proposed the unitary octet model for the symmetry of their strong interactions. He put forward the current algebra which links this unitary symmetry with a chiral

symmetry. Above all, it is Gell-Mann who was responsible for the successful hypothesis that quarks are the subunits from which the mesons and baryons are constructed.

The last section is the least successful. The authors badly needed to consult an active particle physicist before going into print. As it is, the section gives, perhaps, a fair picture of the situation in the field up to about 1975, although various models are rather muddled up together. The

global symmetry model was never a serious contender and scarcely deserves mention, for example. The authors write as if CP (simultaneous particle — antiparticle and space reflections) symmetry were exact, whereas its violation was established for $K_L^0 \rightarrow \pi^+ \pi^-$ decay in 1964. They seem to regard SU(6) symmetry as an established symmetry, which it is not. They often write as if all weak interaction processes involve a neutrino, although they do mention some weak non-leptonic decay processes such as $\Lambda \rightarrow p \pi^-$. Quantum chromodynamics (QCD) is discussed on the last two pages, but there is an error of notation — the gluon states are like $r\bar{r}$, $r\bar{g}$, $r\bar{b}$, $g\bar{r}$, and so on, not like rr , rg , rb , gr . The "30-odd arbitrary parameters" of current particle theory are unfairly

assigned to QCD, whereas QCD, the theory of the gluon field, involves only one parameter, a scale factor. The quark masses do not stem from QCD, but are generally believed to be due to the electroweak interactions, which violate QCD. The recently discovered weak bosons and their relationship with the electroweak interactions are briefly mentioned on the last page. The second paragraph on that page should have been excised — the idea that free quarks may exist with Planck mass values, their binding to other quarks being due to gravitational forces, has no support whatever.

For an encyclopaedic book, the index provided is inadequate. Many scientists who are mentioned in the text, but not in any section title, are not included, and this lowers the value of the book for the serious reader. □

R. H. Dalitz is Professor of Theoretical Physics, Oxford University, Oxford OX1 3NP, UK.

Shining light of physics

N. Straumann

The Collected Papers of Albert Einstein. Volume 2: The Swiss Years: Writings, 1900–1909. Edited by John Stachel. Princeton University Press: 1990. Pp.656. \$85. (English translation by Anna Beck available separately. Pp.398. \$25.)

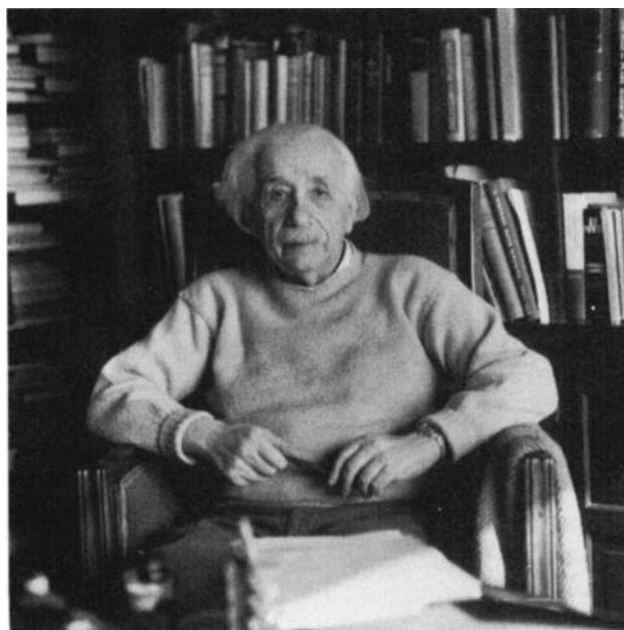
THIS second volume of Einstein's *Collected Papers* might well turn out to be the most important of the whole projected series of approximately

30 volumes. It contains Einstein's scientific work during the first decade of his career, beginning with two papers on the nature of intermolecular forces and ending with his famous Salzburg lecture "On the Development of our Views Concerning the Nature and Constitution of Radiation".

This beautiful volume contains 62 annotated documents presented in facsimile, 8 well-written editorial notes and a useful introduction. By contrast to the first volume, there are no untranslated German quotations in the editorial notes and annotations. An accompanying volume contains the English translations of the documents, all of which were originally written in German. People who know only English should have no problem in reading these in conjunction with the documentary edition and its editorial apparatus.

It is well known that in March, May and June of 1905, Einstein wrote the three masterpieces on three widely different fields of physics which established him as a central figure in twentieth-century physics. But even before that year of miracles, he had published several remarkable papers on the problems of molecular reality and the molecular basis of thermodynamics. In particular, his work on statistical mechanics was absolutely crucial for his most revolutionary contribution to physics: the light-quantum hypothesis. It may be recalled that Einstein extracted the light-quantum postulate from an analogy between radiation in the Wien regime and a classical ideal gas of material particles. In doing so, Boltzmann's principle and Einstein's understanding of statistical fluctuations played an important role.

Einstein wrote three papers devoted exclusively to the foundations of statistical mechanics (documents 3–5), in which he tried to fill what he considered to be a gap in the mechanical foundations of thermodynamics. The harvest of results used in his later work is well described in one of the editorial notes here. When Einstein wrote his three papers he was not familiar with the work of Gibbs (or of some of Boltzmann's). Einstein's papers form a bridge, parallel to that of Gibbs in 1902, between Boltzmann's work and the modern approach to statistical mechanics. In particular, Einstein independently formulated the distinction between the microcanonical and canonical ensembles



Einstein — establishing a central role in modern physics.

and derived the equilibrium distribution for the canonical ensemble from the microcanonical distribution. Very important for his later work was the derivation of the energy fluctuation formula for the canonical ensemble.

In 1907, Einstein wrote a careful review article on relativity for the *Jahrbuch der Radioaktivität und Elektronik* (document 47). In the final section, entitled "Principle of Relativity and Gravitation", Einstein tries to incorporate gravitation into the theory of relativity and arrives at his famous principle of equivalence, from which he deduced the redshift of spectral lines from the Sun and the deflection of light in a gravitational field. Einstein concludes that, if gravitation is to be included, it is necessary to extend the relativity principle. This is the only publication on the problem of gravitation during the period covered by the present volume.

This is not the place to go through many of the documents reproduced in this volume. The last major paper is the published version of an invited talk that Einstein gave to the physics section of the *Gesellschaft Deutscher Naturforscher und*

Aerzte at its 1909 meeting in Salzburg. Pauli once said that this report can be regarded as a turning-point in the development of theoretical physics. Einstein discussed the theory of relativity and quantum theory and pointed out important interconnections between his work on the quantum hypothesis, on relativity, and on brownian motion and statistical physics. In particular, he analysed fluctuations in the energy and momentum of the radiation field, drawing on ideas and methods he had developed earlier in the course of his work on statistical physics and brownian motion. Einstein maintained that "the next phase in the development of theoretical physics will bring us a theory of light that may be conceived as a sort of fusion of the wave and of the emission theory of light". In this talk he gave the first synthesis of the profound transformation in the concept of light brought about by the theory of relativity and the quantum hypothesis, and of the profound implications of this transformation for the development of physics.

Many scientists will read again and again this long-awaited volume, which contains some of the most significant achievements of twentieth-century physics. □

N. Straumann is Professor of Theoretical Physics, University of Zurich, CH-8001 Zurich, Switzerland.

AIDS TARGETED INFORMATION/ATIN

TARGETS YOUR LITERATURE SEARCH

Expand your base of knowledge about AIDS/HIV with *AIDS Targeted Information/ATIN*. Every month, *ATIN* not only provides abstracts but also in-depth evaluations of the current published scientific literature on AIDS. Written by clinicians and researchers for clinicians and researchers, *ATIN* provides an authoritative command of the world's literature on AIDS. Published by Williams & Wilkins. Indexed. 12 issues per year. \$125.

To order call

1-800-638-6423

Sponsored by the American Foundation for AIDS Research (AmFAR).

First lord of British science

David Gooding

Energy and Empire: A Biographical Study of Lord Kelvin. By Crosbie Smith and M. Norton Wise. Cambridge University Press: 1989. Pp. 866. £60, \$89.50.

BORN in Belfast in 1824 and raised and educated in Glasgow, William Thomson made important contributions to field theory, thermodynamics, electrical and navigational instrumentation and to communications, engineering the first successful trans-Atlantic cable in 1866, for which he was knighted that year. He was elevated to the peerage in 1892, the first scientist to be so honoured.

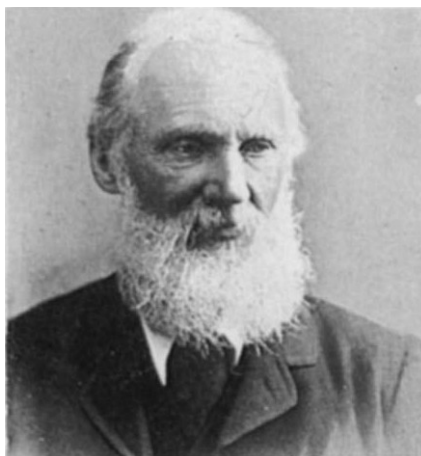
Thomson took the name Kelvin from the river that flows past the University of Glasgow, where he was professor of natural philosophy from 1846 to 1899. Such was his role in creating the new mathematical physics, based on energy, and in developing the technologies of his time, that Smith and Wise have had to write a biography of the family of disciplines and professions that made up Victorian science as well as a life of Thomson. Their book is finely worked, comprehensive, well-written and fully equal to the ambitious pace and grand scale of the life of this first lord of British science. It is a model study of how personal qualities and circumstances interact with social and cultural influences to shape an individual and what he creates.

Educated first by his father and then at Glasgow College and at Cambridge University, William's future was charted by his father, for whom his appointment as professor both secured his son's future and promoted the liberal, democratic and anti-sectarian cause within Glasgow College. Smith and Wise describe the mathematical culture on which Thomson's early papers drew and to which the reforming mathematical style, typified by the new operational calculus he promoted through the *Cambridge and Mathematical Journal*, was a response. (They also rescue the mathematics tutor William Hopkins from undeserved neglect.)

Family correspondence shows that when a Cambridge undergraduate, William remained loyal to the values and ideals of the Thomson family. Having done brilliantly at Cambridge, he took his father's advice to travel to Paris to learn experimental methods in Regnault's laboratory. This ensured that he returned to Glasgow as a practical Scotsman as well as a Cambridge mathematician. Within four years of his appointment as professor at the age of 22, Thomson established the century's most successful applied physics laboratory, remembered for its compasses and precision instruments, but important also for another innovation: using laboratory instruction to teach experimental

practice and mental habits of accuracy and precision. This was Britain's first physics teaching laboratory, harnessing the skills of a large corps of students to produce intellectual capital which he invested in new ventures.

A keen collaborator who shunned disciplinary walls, Thomson had many discussions with Faraday, Maxwell, Joule,



Kelvin — knowledge through measurement.

Rankine, his brother the steam-engineer James Thomson (to name a few) and with P. G. Tait, with whom he wrote the *Treatise on Natural Philosophy*, intended to replace the force-based mechanics of Newton's *Principia* by the new physics of energy. New experimental and mathematical results were as grist to an intellectual engine, driven by the need to minimize waste and extract maximum effect. Smith and Wise document a lifelong concern with efficiency and economy, and Thomson's extraordinary productivity shows how effectively his father taught him to expend his energy in highly efficient ways.

Smith's and Wise's chapters on Thomson's contributions to marine technology and telegraphy and to the mathematical framework that embraced new discoveries in electricity, magnetism and heat are the most definitive yet written. They are a valuable source for those interested in the dynamics of science, engineering and technology, showing how Thomson combined enthusiasm for Fourier's new mathematical methods with respect for Faraday's experimental work (which he had earlier despised), and how he struggled to reconcile the work of Joule and Clayperon *en route* to the second law of thermodynamics. The authors' analysis of conceptual discovery includes the personal circumstances, beliefs and values

that shaped Thomson's perception of wider political and economic processes, so shaping the concepts and methods he would place at the very heart of his science. They demonstrate the combined influence of the different cultures of Cambridge and of Glasgow (though their contrast understates interest in practical, useful science outside England's ancient universities) and show how the Thomson family's Glaswegian perspective repeatedly drew attention to the evils of waste: William and his brother James (apprenticed to Fairbairns at Millwall) noticed water running to waste through canal locks; later they brought attention to the loss of heat in marine steam engines.

From their material environment the two sons drew phenomena expressing the values inculcated by their father. Was William therefore a rude mechanic? Because he rejected Maxwell's liberation of his field equations from the mechanical models that helped articulate them, and because he assiduously paired mathematics and measurement to get to grips with work and waste, it became fashionable to treat Thomson as an object lesson in the dangers of insisting on concrete, personal experience when interpreting concepts, and of emphasizing the commercial exploitation of results. But to dismiss Thomson's science as the physics of the factory and the marketplace is too smug. Thomson never entertained a distinction between theory and practice (such as still informs much science teaching today), so it was natural to identify knowledge with measurement, as when he wrote that "When you can measure what you are speaking about ... you know something about it, and when you cannot measure it ... your knowledge is of a meagre and unsatisfactory kind ..."

Commercial success did not cause Thomson to adopt an economic view of measurement: Smith and Wise show convincingly how economics and epistemology had always been implicitly related. They also link Thomson's pursuit of uniform standards of measurement for the empire with his support of the unionist cause (he opposed the creation of an Irish state independent of Britain) to the liberal, antisectarian attitudes of his Irish origins. This argues a deep and personal connection between the name of the temperature scale and that of the peer, because Kelvin's title was a reward more for his unionist sympathies than for his scientific services to nation and empire.

Smith and Wise attend as much to the personal as to the technical and social dimensions of science, so their account should be read by anyone who wants to understand how people weave the fabric of hard science. □

David Gooding is in the Science Studies Centre of the University of Bath, Claverton Down, Bath BA2 7AY, UK.

Old worlds for new

Len Goodwin

Explorers of the Amazon. By Anthony Smith. Viking: 1990. Pp. 344. £14.99, \$21.95.

Camping with the Prince and Other Tales of Science in Africa. By Thomas A. Bass. Houghton Mifflin: 1990. Pp. 304. \$20.95.

THE Amazon, as well as being the world's greatest river, is also the most mysterious. There was no real map of it until recently — and the map changes as the river floods and tributaries alter courses as they think fit.

Exploration has always been a matter of chance and accident, beginning with the discovery of the coast of Brazil in 1500. Pedro Cabral sailed a bit too far west in search of a good wind to take him around Africa on the way to India. He stayed for 12 days — long enough to say Mass and plant a wooden cross — before continuing on the voyage that, at long last, broke the stranglehold of Venice and the Middle East on trade with India. Anthony Smith's essays tell the stories of the adventurers who uncovered the secrets of thousands of miles of river — fascinating tales based on a wide study of contemporary documents and reports.

The Spanish conquistadors had awesome courage and fortitude, matched only by their appalling cruelty, not only to the hapless Indians, but also to their own countrymen. When nothing was left of the Inca empire, the Spaniards sought further plunder from the mythical El Dorado in the river valleys to the East. One of them, Orellana, having been sent downriver in search of food, was unable (or unwilling) to rejoin his brutal leader, Gonzalo Pizarro, and carried on downstream for 3,000 miles to the Amazon's mouth. Seventeen years later, Lope de Aguirre, having murdered his way to the top, followed Orellana, but only because he intended to use the men under his command to reconquer Peru for himself. Exploration from east to west was by Portuguese traders based in Pará (Belém) at the river's mouth — a quiet operation that, despite the old treaty that gave Spain the lion's share, led to the control of half the continent by Portugal.

Chance governed the first scientific exploration of the river by La Condamine. He was sent to Quito by the French Académie des Sciences in 1735, with permission to make measurements to determine whether the Earth was a sphere swollen at the Equator as proposed by Newton, or swollen at the poles as propounded by Cassini. The matter was

settled fairly rapidly in Newton's favour by a second team sent to Lapland at the same time, but La Condamine stayed in the Andes for eight years, his team's activities being complicated by illness, accident, marriage, madness and murder. He decided to return downriver and, with Maldonado, made the first extensive and reliable scientific observations and collections of the Amazon region.

Fifty years later, good luck enabled one of the most competent scientists of the time, Baron von Humboldt, to explore the river. Friends in Madrid with access to the royal family obtained for him a wide-ranging permit to travel and, quite suddenly, the whole of Spanish America was



Aerial view of a river meandering through the Oriente rainforest of the Amazon.

opened to someone entirely suited to take advantage of a brand new world.

The South American monopolies on quinine and rubber were broken by the clandestine removal of *Cinchona* and *Hevea* seeds and plants by the British to grow in plantations in India and Malaya. But the transport to Kew of thousands of rubber seedlings succeeded only because fraud had led to the loss of the intended cargo of a liner that happened to call — and the hold was empty.

The final, horrific tale in the collection concerns the cruel Arana empire on the Putumayo, tributary to the Amazon, where thousands of Indians were held in slavery to produce rubber to send to the boom town of Manaus. The company nominally in charge was based in London and, with the help of Roger Casement (then British consul-general in Rio, but later hanged for high treason in 1916), an American, Walter Hardenberg, exposed the scandal.

Thomas Bass has written a travel book of quite another kind. An American, hooked on Africa as a teenager in 1968 and feeling the uneasy fit between contemporary politics and traditional life, he returned in 1985 and spent two years accompanying scientific expeditions from Timbuktu to the Zambesi River. He had learned to enjoy travelling with scientists because they "pay attention . . . know the names of things . . . ask intelligent

questions . . . and the best are both critical and self-critical".

The title derives from a rather unhappy visit by the Duke of Edinburgh, as president of the World Wildlife Fund, to a unit studying the complex changes in natural vegetation, crop yields and fishing at the inland delta of the Nile in Mali, on the edge of the Sahara. In Nairobi, Bass accompanied the energetic Kenyan director of the International Centre of Insect Physiology and Ecology in a tour of units seeking to control insect pests by biological methods — predators, parasites and traps — to avoid further damage to the environment by outworn insecticides. He dived into Lake Malawi to observe the breeding

habits of cichlid fishes, surveyed the calamities resulting from the application of European and American farming methods to the thin soils of Africa, and camped with Turkana nomads who, although they know better than anyone how to survive in an unpredictable climate, are now having to deal with the chaos caused by politics and civil warfare, to which they have no life-saving response. Bass also accompanied expeditions to dig up bones and stone implements in the Western Rift valley, and to collect blood from people and animals in a virus survey near the Bussa dam in Nigeria.

Racy, informative, and with much of the text as spoken dialogue, Bass's stories highlight the often disastrous effects of ignoring local knowledge and practices; the continent is "littered with White Elephants — the bleaching bones of animals and aid projects alike". Nevertheless, the author believes in an African counterthrust working against entropic doom and a movement, informed by traditional knowledge and resourcefulness, working for the common good. An interesting, thoughtful book by a sympathetic listener with a feel for the country and its peoples, and the ability, in a swift phrase, to bring up the sights, sounds and smells of Africa. □

Len Goodwin is at Shepperlands Farm, Park Lane, Finchampstead, Berkshire RG11 4QF, UK.

Tilting at windmills

Derek Fordham

Robert E. Peary at the North Pole: A Report to the National Geographic Society. By the Foundation for the Promotion of the Art of Navigation. *The Navigation Foundation, Box 1126, Rockville, Maryland 20850: 1989. Pp. 181 plus appendices. Pbk \$15.*

"I KNOW Admiral Peary reached the Pole. The reason that I know it is that I know Peary", wrote Amundsen. "The Pole at last!" was Peary's claim in 1909, a claim which started sceptics everywhere on attempts to establish that wherever Peary had been, it was not to the Pole that he had, "nailed the stars and stripes".

Wally Herbert, the explorer who would be the first to have reached the Pole by dog sledge if Peary is discredited, was the latest in this line of sceptics to go into print with *The Noose of Laurels* (see my review in *Nature* 339, 435; 1989). This new report, commissioned by the National Geographic Society, seems to be largely a reaction to that book.

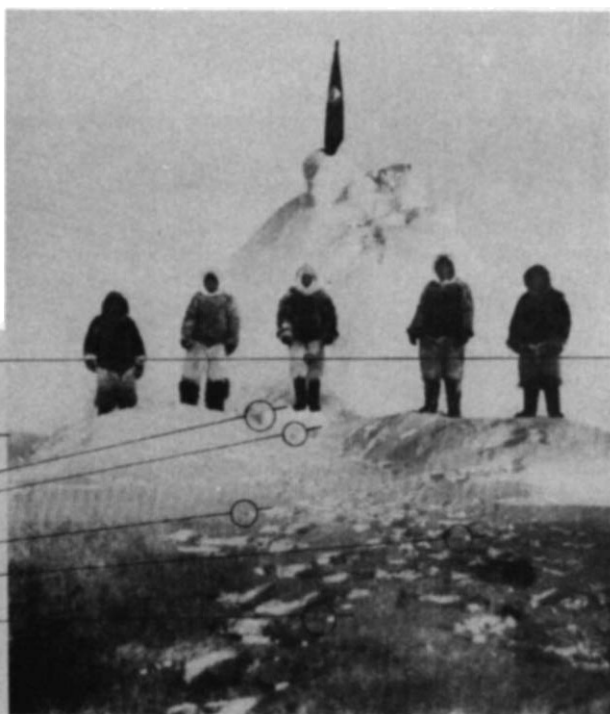
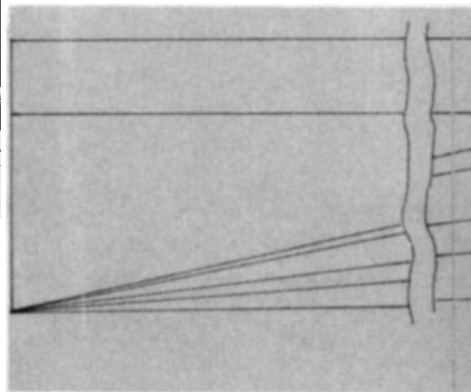
The National Geographic Society was one of Peary's main sponsors and might be expected to find in his favour now just as its three-man committee did after Peary's claim to have reached the Pole in 1909. At times, the Navigation Foundation's apparent determination to refute Herbert's conclusions seems at least as important as the vindication of Peary's case. Their report is just as positive that Peary reached within 5 miles of the Pole as Herbert is that he got no closer than approximately 60 miles.

The authors of this report claim that little new information has been contributed by the succession of books critical of Peary. But despite meticulous research there is also no new information here — new methods of analysis certainly, but the application of state-of-the-art technology to photographic interpretation and Peary's ocean depth soundings in no way overcome many of the objections, particularly those to the claimed daily mileages which form for many polar travellers the core of the objection to Peary's claim. The daily mileages of Will Steger's 1986 North Pole expedition are assumed to support Peary's claims, an assumption all the more misplaced in view of Herbert's careful analysis of the same point.

It is also inappropriate to compare the

daily sledging distances achieved by members of Sverdrup's expedition of 1898 on the relatively smooth sounds and fjords of

Photogrammetrical methods were applied to the shadows seen in Peary's North Pole photographs in an attempt to confirm the geographical location. But uncertainties as to the precise bearing of the Sun and the time of day at which the pictures were taken prevent definite conclusions from being drawn.



the Canadian Arctic archipelago, or those of competitors in the Iditarod race in very different Alaskan conditions, with those possible for Peary on the constantly shifting chaos of the polar pack ice.

Attempts to establish by the application of photogrammetric methods to the shadows on Peary's North Pole photographs that they were taken near the Pole have greater weight, but they suffer from uncertainties as to the precise bearing of the Sun and the time when any of the photographs were taken. These parameters need to be known if the solar angle is to be used to establish a latitude from sight reduction tables.

Nevertheless, the report is a detailed and scholarly summary of probably all there is to know about this Arctic enigma. But it falls short of being a definitive proof, nor does it remove 80 years of doubt, "setting the record straight and putting an end to the long process of vilification of a courageous American explorer", as its authors claim. It sets out a strong case for Peary and in so doing raises some interesting points and throws new light on some of the perceived weaknesses in his claim. Peary's own evidence remains

inconclusive, vague in part and badly presented — the introduction of new technology in support of it merely extends the arena and fuels new possibilities for argument for many years to come. Peary's supporters — and the National Geographic Society is squarely in this camp — will always believe he reached the Pole. It seems likely that attitudes formed over the succeeding 80 long years are too well entrenched to be modified and the situation is best summed up by Franz Werfel:

"For those who believe, no explanation is necessary. For those who do not believe, no explanation is possible." □

Derek Fordham has travelled widely in the Arctic and is at 66 Asburnham Grove, London SE10 8UJ, UK.

■ The natural worlds of the Arctic and Antarctic are the subject of the beautifully illustrated book *Poles Apart*, by Jim Flegg with photographers Eric and David Hosking. The origins, exploration, wildlife, plant life and ecology of both poles are compared. Peary's attempts to reach the North Pole are described, ending with the planting of the "triumphant flag" (page 38) but with no hint of the subsequent controversy described above. Publisher is Pelham, price is £18.99.

■ Also just published, by Blandford, is *The Polar Bear* by Ian Stirling. Again beautifully illustrated with photographs by Dan Guravich, the book is a comprehensive account of the natural history of polar bears, the result of the author's fifteen-year project for the Canadian Wildlife Service. Price is £18.95.

■ *Fur Seal Island* by Paul Thomas will be published on 9 May by Souvenir (£12.95). The author tells an unabashedly emotional story of the northern fur seal, threatened with extinction as a result of human exploitation.

Fruits of the earth

Warwick Bray

The New Archaeology and the Ancient Maya. By Jeremy A. Sabloff. (Scientific American Library). Freeman: 1990. Pp. 193. £16.95, \$32.95.

IN POPULAR imagination (encouraged by the media and the buried treasure industry) an archaeologist is a person who digs things up. Professor Sabloff disposes of this misconception in the opening sentence of chapter 1. "I'm not", he writes, "particularly interested in ancient objects" — and in the rest of the book he enlarges on this statement by explaining, for non-specialists, what archaeologists have really been doing for the past thirty years. Filling museum cases is not the purpose of excavation. Concern has shifted away from artefacts (and from narrative history presented as a sequence of cultures, art styles and dates) to the study of how human societies functioned. Today's archaeologists want to know about institutions, not just about pyramids and potsherds. The search is, first, for patterns of behaviour, next for explanations of these patterns, and then comes the big question: why is it all like this, and what made a particular culture develop as it did? Sabloff is writing about the prehispanic Maya, but the theoretical issues are universal.

In the professional literature, these straightforward aims are all too often concealed behind a screen of jargon and pseudo-philosophy. One of Sabloff's purposes (brilliantly achieved in this book) is to tell the public about the tactics and strategy of archaeological research, not merely about its results, and to do so in language that anybody can understand. He prefers to analyse actual case studies rather than to indulge in philosophical abstraction. The result is a book about archaeology in Maya territory and not, thankfully, another conventional textbook on the ancient Maya.

The author has done plenty of digging in his time, but his full title at the University of Pittsburgh is Professor of Anthropology and History and Philosophy of Science. These are the keywords of the book itself. Sabloff describes the early history of Maya archaeology, with its emphasis on monumental architecture, epigraphy, fine art and the life of the ruling elite. Then, in the early 1960s, came a break with tradition, the "new archaeology" of the title, a movement which had its roots in the philosophy of science (the logical positivism of Carl Hempel and his associates), and which broadened the scope of archaeological enquiry by

introducing new, more rigorous methods of collecting and interpreting data. Investigators of the new generation began to look at the nature of Maya kingship and politics, the relationship between religion and the state, the economics of trade and the control over access to goods, the ecology of subsistence, how people farmed, made tools, disposed of rubbish, and so on.

Having explained the objectives and ideology of the new archaeology, Sabloff examines its effects on Maya studies.

aggressive dynasties. Large urban populations were sustained by intensive agriculture (with terracing and drained fields) rather than by shifting cultivation, and much of the forest that we see today was open farmland during the first millennium AD. Trade products, ideas, marriage partners and, perhaps, armies crossed the internal frontiers of the Maya world, and these connections extended northwards to highland Mexico and south to the Isthmus of Central America. One consequence of recent research is that the Maya have lost



The huge temple-pyramid at Chichén Itzá, once the economic centre of the Yucatán.

Under the onslaught of these new investigations, the traditional view of Maya civilization turned out to be inadequate and often downright nonsense. Sabloff's volume chronicles the rise and fall of a myth, invented by academics but accepted gratefully by a public looking for 'romance' in archaeology. To Sabloff, happiness lies in solving a problem; in other respects he is deliberately anti-romantic, concentrating on new evidence, new ideas and the changes these have brought about in our perception of the Maya.

Most of the things we were told in the 1950s turn out to be untrue. We now know that monumental architecture, complex political organization and the roots of Maya art and religion go back several centuries BC and that the classic Maya were not the peaceful, unworldly intellectuals of the older textbooks. The 'ceremonial centres' of the 1950s model have turned out to be true cities, ruled by

their uniqueness; their culture is no longer an anomaly to be explained away, and in the past twenty years the 'mysterious Maya' have turned into normal Mesoamericans.

This excellent little book tells us what is new, and also how we know these things, what ideas the investigators were testing and what techniques they used. The illustrations, well up to *Scientific American* standards, have been chosen to complement the text, and depict the unglamorous sides of archaeology (surveying, surface-collecting, sieving dirt) as well as the glories of the ancient Maya. In his epilogue, Sabloff notes that the problems of population growth, ecological damage and social stress are not new, and that archaeological studies of past mistakes may help to stop us repeating them. □

Warwick Bray is at the Institute of Archaeology, Gordon Square, London WC1H 0PY, UK.

The Purr-fect Books for Every Cat Lover!



MY LIFE AS A CAT

An Autobiography
by Cornelius The Cat

Cornelius the cat has not only led an interesting life --- He led nine of them! Here in his very first book, *My Life As A Cat -- An Autobiography* Cornelius writes about his nine lives,

beginning as the personal pet of Cleopatra on the Nile up to his final life as a pampered suburban tabby. You'll laugh, you'll cry and you may even fall in love --- with **Cornelius** the cat of nine tales. This highly imaginative book giving you for the first time anywhere a cats eye view of life as a beloved pet. You've never read a book like this before, and we're certain you'll never forget **Cornelius**.

Ever wonder what your cat was actually thinking? Or how it really felt about you? Well, now theres a book that reveals everything! **THE MIND OF A CAT!**, explores the hiterto unexplored world of the feline brain in a simple straight forward style laced with good humor. You'll be amazed by the wealth of scientific insights regarding the inner workings of the cats mind. Yet, you'll be amused endlessly as you discover how a cat percieves the world around him. This unique book is not only rich in gentle humor but filled with impressive, well researched information that will help you appreciate your cat even more. At last, the secrets of the cats mind are revealed - all in this great hardcover book.

THE MIND OF A CAT!

By
BARBARA ARNSTEIN

- ☐ Please send me _____ copies of
MY LIFE AS A CAT - AN AUTOBIOGRAPHY
at \$14.95 (plus \$3.05 postage and handling) - Total \$18.00
☐ **THE MIND OF A CAT!**
at \$14.95 (plus \$3.05 postage and handling) - Total \$18.00
☐ Send me both titles for \$24.95 plus \$3.05 postage and
handling - a savings of \$8.00! - Total \$28.00

GARCO SYSTEMS Dept. D
1405 Avenue Z • Suite 123
Brooklyn, NY 11235

Enclosed find check or money order for \$ _____

Name: _____

Address: _____

City: _____

State/Zip: _____

Only skin deep

Henry Gee

The Dicotylodons: A Study in Palaeobiology. By Gillian King. *Chapman and Hall*: 1990. Pp. 233. £35, \$69.95.

THE Total Perspective Vortex, writes Douglas Adams in *The Hitch-Hikers' Guide to the Galaxy*, is a place where the traveller is granted an instantaneous glimpse of the entire and unimaginable vastness of creation, together with a small plastic sticker that reads 'you are here'. The shock is lethal to anybody with a sense of proportion. The Vortex is constructed on the premise that any event in the Universe, however small, has an effect on everything else. Thus, one might suppose, one could in principle read the complete history of the cosmos from a small piece of fairy-cake.

This idea appeals greatly to palaeontologists, who are on occasion carried away by the urge to re-create entire ecosystems from very small pieces of bone. No wonder that the first view of a fossil site by one raised on big-screen dinoramas can be a chastening experience.

Gillian King is a palaeontologist who habitually carries a hatpin for the expressed purpose of deflating such puffery. *Dicotylodons* is ostensibly a modest work on the evolutionary history of a fossil group. But it contains a hatpin wielded on the very last page, exploding the yarn set out on the previous 199. The effect is appropriately deflating.

The dicotylodons were a group of animals, distantly related to mammals, that lived between about 260 and 230 million years ago. Although possessed of physiognomies of unremitting ugliness, they were the first really successful land herbivores, and have a relatively well-explored fossil record. Dozens of forms are known, mostly from southern Africa. King describes everything that is known about most of the important ones, concen-

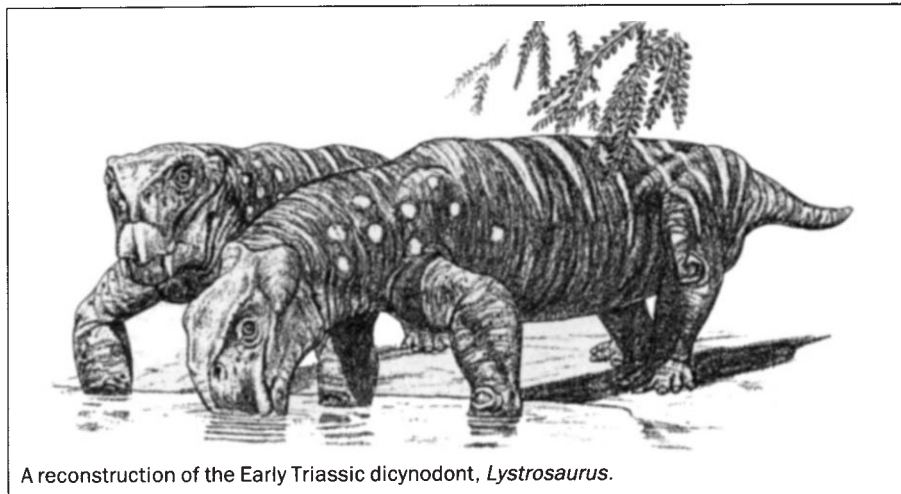
trating on how the reconstructed skulls can be used to infer jaw mechanics and, from this, what the animals ate.

For most of the book, King darts unevenly from topic to topic, veering from rigorous and detailed anatomical work-outs to wide-eyed ecological speculation with jarring effect. But everything becomes clear in the last chapter, in which she discusses what we know about how dicotylodons lived, as real flesh-and-blood animals, for which purpose she perforce descends almost totally into the conditional perfect. Dicotylodons "may have" lived like this, in a world that "may have" been like that, eating plants that "may have" been like something else.

Even though the fossil record of dicotylodons is excellent, this excellence is still only relative: a chilling fact of which King is well aware and is at pains to point out. In the very last paragraph of the book (even the penultimate sentence) she reminds us that "any fossil group keeps far more information to itself than we are ever likely to be able to discover" and then — almost by way of apology — "but that should not stop our trying". The hidden message is that for groups whose fossil records are not even as fine as that of the dicotylodons, ecological reconstructions could all too easily be small triumphs of interpretation over data. Palaeontologists should take note.

My mother once had a book about poultry containing instructions for strangling a chicken. First, it said, one must stretch the neck to its fullest extent. Then stretch it nine inches further. That final nine inches is the measure of the Total Perspective Vortex school of palaeontology, the myth that King explodes, poignantly, with the fossil group of whose interpretation she is one of the world's leading exponents. The book is a lesson to us all, and shows that when faced with the inestimable caprice of geological time, it pays to adopt a sense of proportion. □

Henry Gee is an assistant editor of *Nature*.



A reconstruction of the Early Triassic dicynodont, *Lystrosaurus*.